# Randomized and relaxed strategies in continuous-time Markov decision processes

Alexey Piunovskiy
Department of Mathematical Sciences, University of Liverpool, L69 7ZL, UK.
piunov@liv.ac.uk

**Abstract**

One of the goals of this article is to describe a wide class of control strategies, which includes the traditional relaxed strategies, as well as the so called randomized strategies which appeared earlier only in the framework of semi-Markov decision processes. If the objective is the total expected cost up to the accumulation of jumps, then without loss of generality one can consider only Markov relaxed strategies. Under a simple condition, the Markov randomized strategies are also sufficient. An example shows that the mentioned condition is important. Finally, without any conditions, the class of so called Poisson-related strategies is also sufficient in the optimization problems. All the results are applicable to the discounted model, they may be useful also for the case of long-run average cost.

**Keywords:** continuous-time Markov decision process, total cost, discounted cost, occupation measure, relaxed strategy, randomized strategy
**AMS 2000 subject classification:** Primary 90C40; Secondary 60J25, 60J75.

## 1 Introduction

Continuous-time jump Markov processes, especially Markov chains with the discrete state space $\mathbf{X}$, form a well developed branch of random processes, see, e.g., [2, 24]. After the infinitesimal generator (transition rate) $q(dy|x)$ is fixed, the model is well defined. It can be studied by constructing the canonical sample space and investigating the so called point process; one can directly pass to the transition probability through the Kolmogorov equations. In any case, the model is the same. One can also consider the case of time-dependent transition rate, but in this article we study the homogeneous model.

If we look at the control problem, where the transition rate $q(dy|x,a)$ depends on the action $a$, we face at least two different standard models. If the actions can be changed only at the jump epochs (such actions may also be randomized), then the model is called "Exponential Semi-Markov Decision Process" (ESMDP). If, e.g., two actions $a_1$ and $a_2$ are chosen with probabilities $p(a_1)$ and $p(a_2)=1-p(a_1)$, then the sojourn time in state $x$ has the cumulative distribution function (CDF) $1-\bigl[p(a_1)e^{-q_x(a_1)}+p(a_2)e^{-q_x(a_2)}\bigr]$. Here and below, $q_x(a)$ is the parameter of the exponentially distributed sojourn time in state $x$ under action $a$. The term "Continuous-Time Markov Decision Process" (CTMDP) is for the model where the actions are relaxed: roughly speaking, the actual transition rate at a time moment $t$ is $\int_{\mathbf{A}} q(dy|x,a)\pi(da|t)$, where $\pi(da|\cdot)$ is a predictable process with the values in the space of probability distributions on the action space $\mathbf{A}$. For example, if $\pi(\{a_1\}|t)=\pi(a_1)=1-\pi(a_2)=\pi(\{a_2\}|t)$ then the sojourn time in state $x$ has the CDF $1-e^{-\pi(a_1)q_x(a_1)-\pi(a_2)q_x(a_2)}$. Below, we say "randomized/relaxed strategies", rather than actions. General semi-Markov decision processes, where the sojourn times are not necessarily exponential, were studied in [8, 14, 24], where one can find more relevant references. As soon as the sojourn times are exponential (under a fixed action $a$ and a current state $x$), CTMDP are much more popular: see articles and monographs [7, 9, 10, 11, 16, 20, 23, 25] and references therein. In the case of discounted total expected cost, an excellent discussion of different models can be found in [7]. One of the main results is as follows: for any (relaxed) control strategy in CTMDP, there is

an equivalent (randomized) strategy in ESMDP (and vice versa) meaning that, for any cost rate, the values of the objectives for the corresponding strategies in those two models coincide. In this connection, we have to underline that relaxed strategies are usually not realizable in practice, but randomized strategies can be easily implemented.

In the current article, we use the name CTMDP, but consider a wide class of strategies containing not only any combination of standard relaxations and randomizations (hence covering the traditional CTMDP and ESMDP), but absolutely new strategies like a Brownian motion between the jumps, if the action space is $\mathbf{A} = \mathbb{R}$. To be specific, we investigate the case of the total expected cost, but the developed approach can be useful for other problems, e.g., with the long-run average cost. Note that the discounted cost, including the case of the varying discount factor, is a special case of the total (undiscounted) cost. We allow the transition rate to be non-conservative and arbitrarily unbounded, so that the accumulation of jumps is not excluded.

The main results of the current work are as follows.

– For any control strategy, there is an equivalent Markov purely relaxed strategy (Theorem 2). Here and below, "equivalent" means that the objective values coincide for any given cost rate.

– Under a weak condition, e.g. in the discounted case, for any control strategy, there is an equivalent Markov randomized strategy (Theorem 1) and an equivalent mixture of (simple) deterministic Markov strategies (Theorem 3).

– In general, there can be a relaxed strategy for which no-one randomized strategy is equivalent (Example 2).

– Without any conditions, for any control strategy, there is an equivalent "Poisson-related $\xi$-strategy" (Theorem 5) which is somewhat similar to the so called switching policy [7], but the switching moments as well as the corresponding actions are random. Note, such Poisson-related strategies are easily implementable.

The following remark explains the novelty of the current work and its connection to the previous results and the known methods. As was mentioned (see also Section 5), the discounted cost is a special case of the considered model. Such CTMDP was investigated in [7] where the statements similar to theorems 1 and 2 were proved. Generally speaking, we use the same method of attack, but all the proofs must be carefully rewritten because of the following: a) The occupation measures can take infinite value; b) Markov randomized strategies are not sufficient in optimization problems. The latter is confirmed by Example 2. To cover this gap, we introduce the new sufficient class of Poisson-related $\xi$-strategies.

The CTMDP under study and the control strategies are introduced in Section 2; the main results are formulated in sections 3,4,5 and 6; the proofs are postponed to Appendix. A couple of illustrating examples are given in Section 7.

# 2 Model description

The following notations are frequently used throughout this paper. $\mathbb{N}$ is the set of natural numbers including zero; $\delta_x(\cdot)$ is the Dirac measure concentrated at $x$, we call such distributions degenerate; $I\{\cdot\}$ is the indicator function. $\mathcal{B}(E)$ is the Borel $\sigma$-algebra of the Borel space $E$, $\mathcal{P}(E)$ is the Borel space of probability measures on $E$. $\mathcal{F}_1 \bigvee \mathcal{F}_2$ is the smallest $\sigma$-algebra containing the two $\sigma$-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$. $\mathbb{R}_+ \stackrel{\triangle}{=} (0,\infty)$, $\mathbb{R}_+^0 \stackrel{\triangle}{=} [0,\infty)$, $\bar{\mathbb{R}} = [-\infty,+\infty]$, $\bar{\mathbb{R}}_+ = (0,\infty]$, $\bar{\mathbb{R}}_+^0 = [0,\infty]$. The abbreviation *w.r.t.* (resp. *a.s.*) stands for "with respect to" (resp. "almost surely"); for $b \in \bar{\mathbb{R}}$, $b^+ \stackrel{\triangle}{=} \max\{b,0\}$ and $b^- \stackrel{\triangle}{=} \min\{b,0\}$. If $\mathbf{X}$ and $\mathbf{Y}$ are Borel spaces and $P$ is a probability measure on $\Omega = \mathbf{X} \times \mathbf{Y}$, then, for an integrable function $F(X,Y)$, we denote $E[F(X,Y)|X=x]$ the regular conditional mathematical expectation. In other words, $E[F(X,Y)|X=x]$ is such a measurable function $f$ on $\mathbf{X}$ that $E[F(X,Y)|X] = f(X)$ $P$-a.s. If $\mathbf{Z}$ is an additional Borel space then function $E[F(X,Y,z)|X=x] : \mathbf{X} \times \mathbf{Z} \to \mathbb{R}$ has the same meaning. (This function is measurable [1, Prop.7.29].) Here and usually below, the capital letters denote random variables, and little letters are for their values. The bold letters denote spaces. Equations which involve such conditional expectations, hold a.s. without special remarks.

The primitives of a continuous-time Markov decision process (CTMDP) are the following elements.

- State space: $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ (arbitrary Borel).
- Action space: $(\mathbf{A}, \mathcal{B}(\mathbf{A}))$ (arbitrary Borel), $\mathbf{A}(x) \in \mathcal{B}(\mathbf{A})$ is the non-empty space of admissible actions in state $x \in \mathbf{X}$. It is supposed that $\mathbb{K} \stackrel{\triangle}{=} \{(x,a) \in \mathbf{X} \times \mathbf{A} : \ a \in \mathbf{A}(x)\} \in \mathcal{B}(\mathbf{X} \times \mathbf{A})$ and this set contains the graph of a measurable function from $\mathbf{X}$ to $\mathbf{A}$.
- Transition rate: $q(dy|x,a)$, a signed kernel on $\mathcal{B}(\mathbf{X})$ given $(x,a) \in \mathbb{K}$, taking nonnegative values on $\Gamma_{\mathbf{X}} \setminus \{x\}$ with $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$. We assume that $q(\mathbf{X}|x,a) \le 0$ and $\bar{q}_x \stackrel{\triangle}{=} \sup_{a \in \mathbf{A}(x)} q_x(a) < \infty$, where $q_x(a) \stackrel{\triangle}{=} -q(\{x\}|x,a)$.
- Cost rates: measurable $\bar{\mathbb{R}}$-valued functions $c_i(x,a)$ on $\mathbb{K}$, $i = 0, 1, 2, \ldots, N$.
- Initial distribution: $\gamma(\cdot)$, a probability measure on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$.
- Additional Borel space $(\mathbf{\Xi}, \mathcal{B}(\mathbf{\Xi}))$, the source of the control randomness.

Actually, the space $(\mathbf{\Xi}, \mathcal{B}(\mathbf{\Xi}))$ can be chosen by the decision maker, but it is convenient to introduce it immediately, in order to describe the sample space. The role of the space $\mathbf{\Xi}$ will become clear after the description of control strategies.

We introduce the artificial isolated point (cemetery) $\Delta$, put $\mathbf{X}_\Delta \stackrel{\triangle}{=} \mathbf{X} \cup \{\Delta\}$, $\mathbf{A}_\Delta \stackrel{\triangle}{=} \mathbf{A} \cup \{\Delta\}$, $\mathbf{\Xi}_\Delta = \mathbf{\Xi} \cup \{\Delta\}$, and define $\mathbf{A}(\Delta) \stackrel{\triangle}{=} \Delta$, $q(\Gamma|\Delta,\Delta) \stackrel{\triangle}{=} 0$ for all $\Gamma \in \mathcal{B}(\mathbf{X}_\Delta)$, $\alpha(x,a) \stackrel{\triangle}{=} q(\{\Delta\}|x,a) \stackrel{\triangle}{=} q_x(a) - q(\mathbf{X} \setminus \{x\}|x,a) \ge 0$ for $(x,a) \in \mathbb{K}$. The state $\Delta$ means, the process is over, i.e. escaped from the state space. We also put $c_i(\Delta,\Delta) = 0$.

Given the above primitives, let us construct the underlying (measurable) sample space $(\Omega, \mathcal{F})$. Having firstly defined the measurable space $(\Omega^0, \mathcal{F}^0) \stackrel{\triangle}{=} (\mathbf{\Xi} \times (\mathbf{X} \times \mathbf{\Xi} \times \mathbb{R}_+)^\infty, \mathcal{B}(\mathbf{\Xi} \times (\mathbf{X} \times \mathbf{\Xi} \times \mathbb{R}_+)^\infty))$, let us adjoin all the sequences of the form

$$(\xi_0, x_0, \ \xi_1, \ \theta_1, \ x_1, \ \xi_2, \ \ldots, \ \theta_{m-1}, \ x_{m-1}, \ \xi_m, \ \theta_m, \ \Delta, \ \Delta, \ \infty, \ \Delta, \ \Delta, \ \ldots)$$

to $\Omega^0$, where $m \ge 1$ is some integer, $\xi_m \in \mathbf{\Xi}$, $\theta_m \in \bar{\mathbb{R}}_+$, $\theta_l \in \mathbb{R}_+$ , $x_l \in \mathbf{X}$, $\xi_l \in \mathbf{\Xi}$ for all nonnegative integers $l \le m-1$. After the corresponding modification of the $\sigma$-algebra $\mathcal{F}^0$, we obtain the basic sample space $(\Omega, \mathcal{F})$.

Below,

$$\omega = (\xi_0, \ x_0, \ \xi_1, \ \theta_1, \ x_1, \ \xi_2, \ \theta_2, \ x_2, \ldots).$$

For $n \in \mathbb{N} \setminus \{0\}$, introduce the mapping $\Theta_n : \ \Omega \to \bar{\mathbb{R}}_+$ by $\Theta_n(\omega) = \theta_n$; for $n \in \mathbb{N}$, the mappings $X_n : \ \Omega \to \mathbf{X}_\Delta$ and $\Xi_n : \ \Omega \to \mathbf{\Xi}_\Delta$ are defined by $X_n(\omega) = x_n$ and $\Xi_n(\omega) = \xi_n$. As usual, the argument $\omega$ will be often omitted. The increasing sequence of random variables $T_n$, $n \in \mathbb{N}$ is defined by $T_n = \sum_{i=1}^n \Theta_i$; $T_\infty = \lim_{n\to\infty} T_n$. Here, $\Theta_n$ (resp. $T_n$, $X_n$) can be understood as the sojourn times (resp. the jump moments, the states of the process on the intervals $[T_n, T_{n+1})$). We do not intend to consider the process after $T_\infty$; the isolated point $\Delta$ will be regarded as absorbing; it appears when $\theta_m = \infty$ or when $\theta_m < \infty$ and the jump $x_{m-1} \to \Delta$ is realized with intensity $\alpha(x,a)$. The meaning of the $\xi_n$ components will be described later. Finally, for $n \in \mathbb{N}$,

$$H_n = (\Xi_0, X_0, \Xi_1, \Theta_1, X_1, \ldots, \Xi_n, \Theta_n, X_n)$$

is the $n$-term (random) history. As usual, capital letters $\Xi, X, \Theta, T, H$ denote random elements; the corresponding small letters are for their realizations.

The random measure $\mu$ is a measure on $\mathbb{R}_+ \times \mathbf{\Xi} \times \mathbf{X}_\Delta$ with values in $\mathbb{N} \cup \{\infty\}$, defined by

$$\mu(\omega; \ \Gamma_{\mathbb{R}} \times \Gamma_{\mathbf{\Xi}} \times \Gamma_{\mathbf{X}}) = \sum_{n \ge 1} I\{T_n(\omega) < \infty\} \delta_{(T_n(\omega), \Xi_n(\omega), X_n(\omega))}(\Gamma_{\mathbb{R}} \times \Gamma_{\mathbf{\Xi}} \times \Gamma_{\mathbf{X}});$$

the right continuous filtration $(\mathcal{F})_{t \in \mathbb{R}^0_+}$ on $(\Omega, \mathcal{F})$ is given by

$$\mathcal{F}_t = \sigma\{H_0\} \vee \sigma\{\mu(]0,u] \times B) : \ u \le t, \ B \in \mathcal{B}(\mathbf{\Xi} \times \mathbf{X}_\Delta)\}.$$

The controlled process of our interest

$$X(\omega,t)\stackrel{\triangle}{=}\sum_{n\geq 0} I\{T_n\leq t<T_{n+1}\}X_n+I\{T_\infty\leq t\}\Delta$$

takes values in $\mathbf{X}_\Delta$ and is right continuous and adapted. The filtration $\{\mathcal{F}_t\}_{t\geq 0}$ gives rise to the predictable $\sigma$-algebra on $\Omega\times\mathbb{R}^0_+$ defined by $\mathcal{P}\stackrel{\triangle}{=}\sigma\{\Gamma\times\{0\}\ (\Gamma\in\mathcal{F}_0),\Gamma\times(u,\infty)\ (\Gamma\in\mathcal{F}_{u-},u>0)\}$, where $\mathcal{F}_{u-}\stackrel{\triangle}{=}\bigvee_{t<u}\mathcal{F}_t$. See [16, Chap.4] for more details. $X(t)$ is traditionally called a controlled jump (Markov) process, but in fact, on the constructed sample space, the process $X(t)$ is fixed (not controlled). It will be clear that the probability measure on $(\Omega,\mathcal{F})$ is under control, not the process. Anyway, we will follow the standard terminology.

**Definition 1** *A control strategy is defined as follows*

$$S=\{\mathbf{\Xi},p_0,\langle p_n,\pi_n\rangle,\ n=1,2,\ldots\},$$

*where $p_0(d\xi_0)$ is a probability distribution on $\mathbf{\Xi}$; for $x_{n-1}\in\mathbf{X}$, $p_n(d\xi_n|h_{n-1})$ is a stochastic kernel on $\mathbf{\Xi}$ given $\mathbf{H}_{n-1}$ (the space of $(n-1)$-component histories); $\pi_n(da|h_{n-1},\xi_n,u)$ is a stochastic kernel on $\mathbf{A}(x_{n-1})$ given $\mathbf{H}_{n-1}\times\mathbf{\Xi}\times\mathbb{R}_+$. If $x_{n-1}=\Delta$, then we assume that $p_n(d\xi_n|h_{n-1})=\delta_\Delta(d\xi_n)$ and $\pi_n(da|h_{n-1},\Delta,u)=\delta_\Delta(da)$.*

*A strategy will be called quasi-stationary if the stochastic kernels $p(d\xi_n|\xi_0,x_{n-1})$ and $\pi(da|\xi_0,x_{n-1},\xi_n,u)$ depend on the shown arguments only.*

The $p_n$ components mean the randomizations of controls; the $\pi_n$ components mean relaxations. Below, for $\Gamma_{\mathbf{A}}\in\mathcal{B}(\mathbf{A}_\Delta)$, $t\in\mathbb{R}_+$,

$$\pi(\Gamma_{\mathbf{A}}|\omega,t)=\sum_{n\geq 1} I\{T_{n-1}<t\leq T_n\}\pi_n(\Gamma_{\mathbf{A}}|H_{n-1},\Xi_n,t-T_{n-1});$$

the argument $\omega$ is often omitted.

If the randomizations are absent, that is, the kernels $\pi_n$ do not depend on the $\xi$-components, then we deal with a relaxed strategy. One can omit the $\xi_n$ components; as a result we obtain the standard control strategy $\{\pi_n,\ n=1,2,\ldots\}$; in this case the stochastic kernel

$$\pi(\Gamma_{\mathbf{A}}|\omega,t)=\sum_{n\geq 1} I\{T_{n-1}<t\leq T_n)\pi_n(\Gamma_{\mathbf{A}}|X_0,\Theta_1,\ldots,X_{n-1},t-T_{n-1})$$

is predictable. (This reasoning holds also if the kernels $\pi_n$ depend only on $\xi_0$.) Such models were built and investigated by many authors [7, 9, 10, 11, 16, 20, 23, 25]. Note that the realizations of a relaxed strategy are usually impossible on practice, unless all the transition probabilities $\pi_n$ are degenerate, i.e. are concentrated at singletons

$$\varphi_n(x_0,\theta_1,\ldots,x_{n-1},u)\in\mathbf{A}(x_{n-1}). \tag{1}$$

For a discussion, see [7, p.509]: if, e.g. $\pi_n(\{a_1\}|x_0,\theta_1,\ldots,x_{n-1},u)=\pi_n(\{a_2\}|x_0,\theta_1,\ldots,x_{n-1},u)=0.5$ then the decision maker intends to use the actions $a_1$ and $a_2$ equiprobably at each time moment, but in this case the trajectories of the action process are not measurable.

On the other hand, if the relaxations are absent, that is, all kernels $\pi_n$ are degenerate and are described by measurable functions $\varphi_n$ like in (1), then the action (or control) process $A(t)$ can be defined like follows

$$A(\omega,t)=\sum_{n\geq 1} I\{T_{n-1}<t\leq T_n\}\varphi_n(\Xi_0,X_0,\Xi_1,\Theta_1,\ldots,X_{n-1},\Xi_n,t-T_{n-1})+I\{T_\infty\leq t\}\Delta. \tag{2}$$

Clearly, the $A(t)$ process is measurable, but not necessarily predictable or even adapted. Below, we call such (purely randomized) strategies as $\xi$-strategies; they are defined by sequences $\{\mathbf{\Xi},p_0,\langle p_n,\varphi_n\rangle,\ n=1,2,\ldots\}$. According to (2), after the history $H_{n-1}$ is realized, the decision maker flips a coin resulting in the value of $\Xi_n$ having the distribution $p_n$. Afterwards, up to the next jump epoch $T_n$, the control $A(t)$ is just a (deterministic measurable) function $\varphi_n$.

**Definition 2** *$\xi$-strategies were defined just above. Purely relaxed strategies introduced earlier will be called $\pi$-strategies. General strategies $S$ can be called $\pi$-$\xi$-strategies. If $\pi_n(da|x_0,\theta_1,x_1,\theta_2,\ldots,x_{n-1},u)=\pi_n^M(da|x_{n-1},u)$ for all $n=1,2,\ldots$ then the $\pi$-strategy is called Markov. It is called stationary if $\pi_n^M(da|x_{n-1},u)\equiv\pi(da|x_{n-1})$.*

Suppose a $\pi$-$\xi$-strategy $S$ is fixed. The dynamics of the controlled process can be described like follows. First of all, $\Xi_0=\xi_0$ is realized based on the chosen distribution $p_0(d\xi_0)$. Recall that the realized values of random elements are denoted with the corresponding small letters. If $p_0$ is a combination of two Dirac measures, then in the future this or that control will be applied: $p_0$ is responsible for the mixtures of simpler control strategies. After that, the initial state $X_0$, having the distribution $\gamma(dx)$, is realized. Later, when the realized state $x_{n-1}\in\mathbf{X}$ becomes known at the realized jump epoch $t_{n-1}$ $(n=1,2,\ldots)$, the dynamics is controlled in the following way. The decision maker flips a coin resulting in the $\Xi_n=\xi_n$ component having distribution $p_n(d\xi_n|h_{n-1})$; after that the stochastic kernel $\pi_n(da|h_{n-1},\xi_n,u)$ gives rise to the jumps intensity $\lambda_n(\Gamma|h_{n-1},u)$ from the current state $x_{n-1}$ to $\Gamma\in\mathcal{B}(\mathbf{X}_\Delta)$, where

$$\lambda_n(\Gamma|h_{n-1},\xi_n,u)=\int_{\mathbf{A}}\pi_n(da|h_{n-1},\xi_n,u)q(\Gamma\setminus\{x_{n-1}\}|x_{n-1},a); \tag{3}$$

parameter $u>0$ is the time interval passed after the jump epoch $t_{n-1}$. After the corresponding interval $\theta_n$, the new state $x_n\in\mathbf{X}_\Delta$ of the process $X(t)$ is realized at the jump epoch $t_n=t_{n-1}+\theta_n$. The joint distribution of $(\Theta_n,X_n)$ is given below. And so on. If $\theta_n=\infty$ then $x_n=\Delta$ and actually the process is over: the triples $(\theta=\infty,\Delta,\Delta)$ will be repeated endlessly. The same happens if $\theta_n<\infty$ and $x_n=\Delta$. Along with the intensity $\lambda_n$, we need the following integral

$$\Lambda_n(\Gamma,h_{n-1},\xi_n,t)=\int_{(0,t]\cap\mathbb{R}_+}\lambda_n(\Gamma|h_{n-1},\xi_n,u)du. \tag{4}$$

Note that, in case $q_x(a)\geq\varepsilon>0$, $\Lambda_n(\mathbf{X}_\Delta|h_{n-1},\xi_n,\infty)=\infty$ if $x_{n-1}\neq\Delta$.

Now, the distribution of $H_0=(\Xi_0,X_0)$ is given by $p_0(d\xi_0)\cdot\gamma(dx_0)$ and, for any $n\in\mathbb{N}\setminus\{0\}$, the stochastic kernel $G_n$ on $\bar{\mathbb{R}}_+\times\mathbf{\Xi}_\Delta\times\mathbf{X}_\Delta$ given $\mathbf{H}_{n-1}$ is defined by formulae

$$\begin{aligned}
G_n(\{\infty\}\times\{\Delta\}\times\{\Delta\}|h_{n-1}) &= \delta_{x_{n-1}}(\{\Delta\});\\
G_n(\{\infty\}\times\Gamma_{\mathbf{\Xi}}\times\{\Delta\}|h_{n-1}) &= \delta_{x_{n-1}}(\mathbf{X})\int_{\Gamma_{\mathbf{\Xi}}}e^{-\Lambda(\mathbf{X}_\Delta,h_{n-1},\xi_n,\infty)}p_n(d\xi_n|h_{n-1});\\
G_n(\Gamma_{\mathbb{R}}\times\Gamma_{\mathbf{\Xi}}\times\Gamma_{\mathbf{X}}|h_{n-1}) &= \delta_{x_{n-1}}(\mathbf{X})\int_{\Gamma_{\mathbf{\Xi}}}\int_{\Gamma_{\mathbb{R}}}\lambda_n(\Gamma_{\mathbf{X}}|h_{n-1},\xi_n,t)\\
&\quad\times e^{-\Lambda_n(\mathbf{X}_\Delta,h_{n-1},\xi_n,t)}dt\;p_n(d\xi_n|h_{n-1});\\
G_n(\{\infty\}\times\mathbf{\Xi}_\Delta\times\mathbf{X}|h_{n-1}) &= G_n(\mathbb{R}_+\times\{\Delta\}\times\mathbf{X}_\Delta|h_{n-1})=0.
\end{aligned} \tag{5}$$

Here $\Gamma_{\mathbb{R}}\in\mathcal{B}(\mathbb{R}_+)$, $\Gamma_{\mathbf{\Xi}}\in\mathcal{B}(\mathbf{\Xi})$, $\Gamma_{\mathbf{X}}\in\mathcal{B}(\mathbf{X}_\Delta)$.

It remains to apply the induction and Ionescu-Tulcea's theorem [1, Prop.7.28] or [18, p.294] to obtain the probability measure $P_\gamma^S$ on $(\Omega,\mathcal{F})$ called strategic measure. According to [15, Prop.3.1], the following formula defines a version of the predictable projection of $\mu$, again a measure on $\mathbb{R}_+\times\mathbf{\Xi}\times\mathbf{X}_\Delta$

$$\begin{aligned}
\nu(\omega;dt,d\xi,dx) &= \sum_{n\geq1}\frac{G_n(dt-T_{n-1},d\xi,dx|H_{n-1})}{G_n([t-T_{n-1},\infty]\times\mathbf{\Xi}_\Delta\times\mathbf{X}_\Delta|H_{n-1})}I\{T_{n-1}<t\leq T_n\}\\
&= \sum_{n\geq1}\frac{p_n(d\xi|H_{n-1})\lambda_n(dx|H_{n-1},\xi,t-T_{n-1})e^{-\Lambda_n(\mathbf{X},H_{n-1},\xi,t-T_{n-1})}}{\int_{\mathbf{\Xi}}e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\xi,t-T_{n-1})}p_n(d\xi|H_{n-1})}dt\;I\{T_{n-1}<t\leq T_n\}.
\end{aligned}$$

Below, when $\gamma(\cdot)$ is a Dirac measure concentrated at $x\in\mathbf{X}$, we use the 'degenerated' notation $P_x^S$. Expectations with respect to $P_\gamma^S$ and $P_x^S$ are denoted as $E_\gamma^S$ and $E_x^S$, respectively. The set of all $\pi$-$\xi$-strategies $S$ will be denoted as $\Pi_S$; the collections of all $\pi$- and $\xi$-strategies will be denoted as $\Pi_\pi$ and $\Pi_\xi$ correspondingly.

We aim to study several classes of control strategies and the associated measures. That is important for stochastic optimal control. For example, one can consider the following two specific problems.

1. Unconstrained problem.

$$\begin{aligned} W_0(S) &= E_\gamma^S\left[\sum_{n=1}^{\infty}\int_{(T_{n-1},T_n]}\int_{\mathbf{A}}\pi_n(da|H_{n-1},\Xi_n,t-T_{n-1})c_0^+(X_{n-1},a)dt\right] \\ &\quad +E_\gamma^S\left[\sum_{n=1}^{\infty}\int_{(T_{n-1},T_n]}\int_{\mathbf{A}}\pi_n(da|H_{n-1},\Xi_n,t-T_{n-1})c_0^-(X_{n-1},a)dt\right] \qquad (6) \\ &= E_\gamma^S\left[\int_{(0,T_\infty)}\int_{\mathbf{A}}\pi(da|t)c_0(X(t),a)\ dt\right] \to \inf_{S\in\Pi_S}. \end{aligned}$$

Here and below, $\infty-\infty\stackrel{\triangle}{=}+\infty$.

2. Constrained problem.

$$\left.\begin{array}{ll} & W_0(S)\to\inf\limits_{S\in\Pi_S} \\ \text{subject to} & \\ & W_i(S)\le d_i,\quad i=1,2,\ldots,N, \end{array}\right\} \qquad (7)$$

where all the objectives $W_i(S)$ have the form similar to (6) with function $c_0$ being replaced with other given cost rates $c_i(x,a)$; $d_i$ are given numbers. All mathematical expectations and integrals of a real function $r$ are calculated separately for $r^+$ and $r^-$ as was demonstrated in (6). As usual, a strategy $S^*$ is called optimal ($\delta$-optimal) in the problem (6) or (7) if $W_0(S^*)$ provides the infimum (is in the $\delta$-neighbourhood of the infimum) and satisfies all the constraints.

The results presented in the current article are also useful for other (constrained) optimal control problems: see the remark after Theorem 2.

**Remark 1** *Suppose a strategy $S$ is such that, for some $m\ge 0$, all kernels $\{\pi_n\}_{n=1}^\infty$ for $x_{n-1}\ne\Delta$ do not depend on the $\xi_m$-component. Then one can omit $\xi_m\in\mathbf{\Xi}_\Delta$ and $\Xi_m\in\mathbf{\Xi}_\Delta$ from the consideration. In this case, instead of the strategic measure $P_\gamma^S(d\omega)$, we can everywhere use the marginal $\tilde{P}_\gamma^S(d\tilde{\omega})=P_\gamma^S(d\tilde{\omega}\times\mathbf{\Xi})$. Here*

$$\tilde{\omega}=(\xi_0,x_0,\xi_1,\theta_1,\ldots,x_{m-1},\theta_m,x_m,\xi_{m+1},\theta_{m+1},\ldots)$$

*and $\tilde{\omega}\times\mathbf{\Xi}=(\xi_0,x_0,\xi_1,\theta_1,\ldots,x_{m-1},\mathbf{\Xi},\theta_m,x_m,\xi_{m+1},\theta_{m+1},\ldots)$. Below, we omit the tilde and hope this will not lead to a confusion.*

*For example, for a purely relaxed strategy $S\in\Pi_\pi$, the strategic measure is defined on the space of sequences*

$$\omega=(x_0,\theta_1,x_1,\ldots).$$

*Another important case is when only the $\xi_0$-component plays a role; then $\omega=(\xi_0,x_0,\theta_1,x_1,\ldots)$ and such a strategy is a mixture of (relaxed) strategies. More about mixtures in Definition 5 and in Section 4.*

**Definition 3** *Purely deterministic strategies, when the functions $\varphi_n$ in (2) do not depend on the $\xi$-components, can be equally called $\pi$-strategies (with degenerate kernels $\pi_n$) or $\xi$-strategies; they are defined by sequences $\{\varphi_n,\ n=1,2,\ldots\}$; the $\xi$-components are omitted. We always assume that $\varphi_n(h_{n-1},u)=\Delta$ if $x_{n-1}=\Delta$. A deterministic Markov strategy is defined by the mappings $\{\varphi_n(x_{n-1},u),\ n=1,2,\ldots\}$. If the mappings $\varphi_n(x_{n-1},u)=\hat{\varphi}_n(x_{n-1})$ do not depend on $u$, the strategy is called simple deterministic Markov. A stationary deterministic strategy is defined by a function $\varphi^s(x)$.*

In case the mappings $\hat{\varphi}_n(\xi_0,x_{n-1})$ depend additionally on the $\xi_0$-component, the strategy will be called a mixture of simple deterministic Markov strategies. A little more general construction is given below: see Definition 5.

As was mentioned, the space $\mathbf{\Xi}$ can be chosen by the decision maker. Let us look at several possibilities.

**Definition 4** *Suppose $\Xi = \mathbf{A}$ and the relaxations are absent, i.e. we deal with a $\xi$-strategy, and the functions $\varphi_n$ in (2) have the form $\varphi_n(h_{n-1}, \xi_n, u) = \xi_n$, so that the argument $\xi_0$ never appears and thus can be omitted. Then such a strategy will be called a standard $\xi$-strategy. It will be denoted as $S = \{\mathbf{A},\ p_n,\ n = 1, 2, \ldots\}$ and below we usually write $A_n$ (or $a_n$) instead of $\Xi_n$ (or $\xi_n$), $n = 1, 2, \ldots$. If we consider only such strategies then we deal with the so called ESMDP [7, p.498]. In case $p_n(d\xi_n|h_{n-1}) = p_n(da_n|h_{n-1}) = p_n^M(da_n|x_{n-1})$ $(n = 1, 2, \ldots)$, the standard $\xi$-strategy will be called Markov; it will be called stationary if the kernels $p_n(da_n|h_{n-1}) = p^s(da_n|x_{n-1})$ do not depend on $n$. A Markov standard $\xi$-strategy with the degenerate kernels $p_n^M(da_n|x_{n-1}) = \delta_{\hat{\varphi}_n(x_{n-1})}(da_n)$, $n = 1, 2, \ldots$ is obviously simple deterministic Markov. The collection of all Markov (stationary) standard $\xi$-strategies will be denoted as $\Pi_\xi^M$ $(\Pi_\xi^s)$, they are often denoted as $p^m$ and $p^s$ instead of $S$, correspondingly.*

Another meaningful case corresponds to the Skorohod space $\Xi = D_{\mathbf{A}}[0, \infty)$, the space of right continuous $\mathbf{A}$-valued functions of time with left limits, endowed with the Skorohod metric [5, Ch.3,§5]. Here we assume that the metric in $\mathbf{A}$ is fixed, such that $\mathbf{A}$ is a Polish space (separable and complete). Now, $D_{\mathbf{A}}[0, \infty)$ is again a Polish space [5, Ch.3,Th.5.6] and hence Borel. Again suppose the relaxations are absent, i.e. consider a $\xi$-strategy, and put

$$\varphi_n(\xi_0, x_0, \xi_1, \theta_1, \ldots, x_{n-1}, \xi_n, u) = \xi_n(u).$$

**Lemma 1** *The mapping $(\xi_n, u) \to \xi_n(u)$ is measurable.*

The proofs of this and other statements are given in Appendix.

Now it is clear that the action (control) process $A(t)$ given by (2) is well defined (that is, measurable) for any $\xi$-strategy. For example, if $A = (-\infty, +\infty)$ then, under appropriately chosen distributions $p_n$, the $A(t)$ process may be a Brownian motion. Such possibilities were never considered before.

**Definition 5** *Consider a $\xi$-strategy $S = \{\Xi,\ p_0,\ \langle p_n, \varphi_n\rangle,\ n = 1, 2, \ldots\}$ satisfying the following conditions: $\Xi = \Xi^{\mathbf{0}} \times \mathbf{A}$, so that $\xi = (\xi^0, a)$; the stochastic kernels*

$$p_n(d\xi_n^0, da_n|\xi_0^0, x_0, a_1, \theta_1, x_1, a_2, \ldots, \theta_{n-1}, x_{n-1})$$

*depend only on the shown components, and $\varphi_n(h_{n-1}, (\xi_n^0, a_n), u) = a_n$. We call $S$ a mixture of standard $\xi$-strategies*

$$S^{\xi_0^0} = \{\mathbf{A}, \hat{p}_n(da|\xi_0^0, x_0, a_1, \theta_1, \ldots, x_{n-1}) \stackrel{\triangle}{=} p_n(\Xi^0 \times da|\xi_0^0, x_0, a_1, \theta_1, \ldots, x_{n-1}),\ n = 1, 2, \ldots\}.$$

*The elements $a_0$ and $\xi_n^0$, $n = 1, 2, \ldots$ play no role, and we omit them. (See Remark 1.) Since only the marginal distributions $\hat{p}_0(d\xi_0^0) = p_0(d\xi_0^0 \times \mathbf{A})$ and $\hat{p}_n(da_n|\xi_0^0, x_0, a_1, \theta_1, \ldots, x_{n-1})$ are important, we denote such a mixture as $\{\Xi^{\mathbf{0}} \times \mathbf{A},\ \hat{p}_0,\ \hat{p}_n,\ n = 1, 2, \ldots\}$.*

*We call $S$ a mixture of simple deterministic Markov strategies $S^{\xi_0^0} = \{\hat{\varphi}_n^{\xi_0^0},\ n = 1, 2, \ldots\}$ in case $\forall \xi_0^0 \in \Xi^0$*

$$\hat{p}_n(\Gamma_{\mathbf{A}}|\xi_0^0, X_0, A_1, \Theta_1, \ldots, X_{n-1}) = I\{\Gamma_{\mathbf{A}} \ni \hat{\varphi}_n^{\xi_0^0}(X_{n-1})\} \quad P_\gamma^S\text{-a.s.} \quad n = 1, 2, \ldots\ ,$$

*where $\{\hat{\varphi}_n^{\xi_0^0},\ n = 1, 2, \ldots\}$ is a simple deterministic Markov strategy. Note, we do not require $\hat{\varphi}_n^{\xi_0^0}(x)$ to be $\Xi^0 \times \mathbf{X}$-measurable. More about such mixtures in Section 4.*

According to the definitions, the intersection of $\xi$-strategies and $\pi$-strategies coincides with the set of purely deterministic strategies. Its subset, the class of stationary deterministic strategies, is the intersection of stationary $\pi$-strategies and $\xi$-strategies. This class is a subset of simple deterministic Markov $\xi$-strategies, and also a subset of stationary standard $\xi$-strategies. Under the compactness-continuity conditions, this set is sufficient for solving many specific single-objective optimal control problems [10, 23]. One can easily establish other relations between the introduced classes of strategies. Note that a mixture of standard $\xi$-strategies is not a $\pi$-strategy.

Let us remind that, if we consider only standard $\xi$-strategies, then in fact we deal with ESMDP. On the other hand, if we consider only $\pi$-strategies, then we are in the framework of traditional CTMDP.

According to Remark 1, slightly modified sample spaces are associated with different types of strategies which are again denoted in different ways. For the reader's convenience, we summarize the main notations in Table 1.

Table 1:

| Strategy | Sample space |
|---|---|
| General ($\pi$-$\xi$-strategy)<br>$S = \{\mathbf{\Xi},\ p_0,\ \langle p_n, \pi_n\rangle,\ n = 1, 2, \ldots\} \in \Pi_S$ | $\Omega = \{(\xi_0, x_0, \xi_1, \theta_1, x_1, \xi_2, \theta_2, \ldots)\}$ |
| Purely randomized ($\xi$-strategy)<br>$S = \{\mathbf{\Xi},\ p_0,\ \langle p_n, \varphi_n\rangle,\ n = 1, 2, \ldots\} \in \Pi_\xi$ | $\Omega = \{(\xi_0, x_0, \xi_1, \theta_1, x_1, \xi_2, \theta_2, \ldots)\}$ |
| Purely relaxed ($\pi$-strategy)<br>$S = \{\pi_n,\ n = 1, 2, \ldots\} \in \Pi_\pi$ | $\Omega = \{(x_0, \theta_1, x_1, \theta_2, \ldots)\}$ |
| Purely deterministic<br>$S = \{\varphi_n(x_0, \theta_1, \ldots, x_{n-1}, s),\ n = 1, 2, \ldots\}$ | $\Omega = \{(x_0, \theta_1, x_1, \theta_2, \ldots)\}$ |
| Simple deterministic Markov<br>$S = \{\hat{\varphi}_n(x_{n-1}),\ n = 1, 2, \ldots\}$ | $\Omega = \{(x_0, \theta_1, x_1, \theta_2, \ldots)\}$ |
| Standard $\xi$-strategy<br>$S = \{\mathbf{A},\ p_n(h_{n-1}),\ n = 1, 2, \ldots\}$ | $\Omega = \{(x_0, \xi_1 = a_1, \theta_1, x_1, \xi_2 = a_2, \theta_2, \ldots)\}$ |
| Markov standard $\xi$-strategy<br>$S = \{\mathbf{A},\ p_n^M(da_n\|x_{n-1}),\ n = 1, 2, \ldots\} = p^M \in \Pi_\xi^M$ | $\Omega = \{(x_0, \xi_1 = a_1, \theta_1, x_1, \xi_2 = a_2, \theta_2, \ldots)\}$ |
| Stationary standard $\xi$-strategy<br>$S = \{\mathbf{A},\ p^s(da\|x)\} = p^s \in \Pi_\xi^s$ | $\Omega = \{(x_0, \xi_1 = a_1, \theta_1, x_1, \xi_2 = a_2, \theta_2, \ldots)\}$ |
| Mixture of standard $\xi$-strategies<br>$\{\mathbf{\Xi^0} \times \mathbf{A},\ \hat{p}_0(d\xi_0^0),\ \hat{p}_n(da_n\|h_{n-1}),\ n = 1, 2, \ldots\}$ | $\Omega = \{(\xi_0, x_0, a_1, \theta_1, x_1, a_2, \theta_2, \ldots)\}$ |

We introduced the new, more rich set of strategies $\Pi_S$, and one of the targets is to establish the sufficiency of smaller classes ($\pi$-strategies, $\xi$-strategies, mixtures, and so on).

# 3 Occupation measures and sufficient classes of strategies

**Definition 6** *For a fixed strategy $S \in \Pi_S$, we introduce the occupation measures*

$$\eta_n^S(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = E_\gamma^S \left[ \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | H_{n-1}, \Xi_n, t - T_{n-1}) dt \right], \quad n = 1, 2, \ldots,$$

*where $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X}), \Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A})$. Note, measure $\eta_n^S$ may be not finite, e.g. if $\Theta_n = \infty$.*

*If $S$ is a standard $\xi$-strategy, or a mixture of standard $\xi$-strategies, then*

$$\eta_n^S(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = E_\gamma^S \left[ I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} I\{A_n \in \Gamma_{\mathbf{A}}\} \Theta_n \right], \quad n = 1, 2, \ldots$$

For any non-negative function $r(x, a)$, for any $S \in \Pi_S$,

$$E_\gamma^S \left[ \sum_{n=1}^{\infty} \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} \int_{\mathbf{A}} \pi_n(da | H_{n-1}, \Xi_n, t - T_{n-1}) r(X_{n-1}, a) dt \right] = \sum_{n=1}^{\infty} \int_{\mathbf{X} \times \mathbf{A}} r(x, a) \eta_n^S(dx, da). \tag{8}$$

In the previous expressions, one can write open intervals $(T_{n-1}, T_n)$, leading to the same occupation measures and cost functionals.

Now, after we introduce the sets $\mathcal{D}_S = \{\ \{\eta_n^S\}_{n=1}^\infty,\ S \in \Pi_S\}$, $\mathcal{D}_\pi = \{\ \{\eta_n^S\}_{n=1}^\infty,\ S \in \Pi_\pi,$ $S$ is Markov$\}$ and $\mathcal{D}_\xi = \{\ \{\eta_n^S\}_{n=1}^\infty,\ S \in \Pi_\xi$ with $\mathbf{\Xi} = \mathbf{A}$, $\xi$-strategy $S$ is Markov standard$\}$, the problems (6) and (7) can be reformulated as

$$\sum_{n=1}^\infty \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\eta_n(dx,da) \to \inf_{\{\eta_n\}_{n=1}^\infty \in \mathcal{D}_S}$$

and

$$\left.\begin{array}{ll} & \displaystyle\sum_{n=1}^\infty \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\eta_n(dx,da) \to \inf_{\{\eta_n\}_{n=1}^\infty \in \mathcal{D}_S} \\ \text{subject to} & \\ & \displaystyle\sum_{n=1}^\infty \int_{\mathbf{X}\times\mathbf{A}} c_i(x,a)\eta_n(dx,da) \le d_i, \quad i = 1,2,\dots,N, \end{array}\right\}$$

correspondingly.

**Condition 1** *(a)* $q_x(a) > 0$ *for all* $(x,a) \in \mathbb{K}$.
*(b)* $\exists \varepsilon > 0 :\ \forall x \in \mathbf{X}\ \inf_{a\in\mathbf{A}(x)} q_x(a) \ge \varepsilon$.

As explained in Section 5, the classical discounted model satisfies the requirement 1-(b). Certainly, if $q_x(a) = 0$ for some $(x,a) \in \mathbb{K}$, and that state $x$ cannot be reached under any control strategy $S$, then one can consider the state space $\mathbf{X} \setminus \{x\}$. Similarly, if $q_x(a) \equiv 0$ for all $a \in \mathbf{A}(x)$ and $\forall i = 0,1,2,\dots,N$, $\forall n = 1,2,\dots$ $c_i(x,a) \equiv 0$ for all $a \in \mathbf{A}(x)$, then one can denote that state $x$ as $\Delta$ (meaning, the process escaped from the state space $\mathbf{X}$). The situation, when $q_x(a) = 0$ and $c_i(x,a) \ne 0$ for a reachable state $x$ and for some $i$ and $a \in \mathbf{A}(x)$, is more delicate.

**Theorem 1** *Suppose Condition 1-(a) is satisfied. Then, for any* $\pi$-$\xi$*-strategy* $S$*, there is a Markov standard* $\xi$*-strategy* $S_\xi$ *such that* $\eta_n^{S_\xi} \ge \eta_n^S$ *for all* $n = 1,2,\dots$. *Hence, Markov standard* $\xi$*-strategies are sufficient for solving optimization problems (6) and (7) with negative costs* $c_i$.

*If Condition 1-(b) is satisfied, then* $\mathcal{D}_S = \mathcal{D}_\xi$. *Hence, Markov standard* $\xi$*-strategies are sufficient in the problems (6) and (7).*

It follows from the proof given in Appendix that one can slightly weaken Condition 1-(b): $\mathcal{D}_S = \mathcal{D}_\xi$ if, for any control strategy $S$,

$$\delta_{X_{n-1}}(\mathbf{X}) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, \infty)} = 0 \quad P_\gamma^S\text{-a.s. for all } n = 1,2,\dots \tag{9}$$

Besides, if a particular $\pi$-$\xi$-strategy $S$ is such that equality (9) is valid, then there is a Markov standard $\xi$-strategy $S_\xi$ such that $\eta_n^{S_\xi} = \eta_n^S$ for all $n = 1,2,\dots$ .

**Corollary 1** *All the statements of Theorem 1 remain valid if we consider only quasi-stationary* $\pi$-$\xi$*-strategies and stationary standard* $\xi$*-strategies.*

**Theorem 2** $\mathcal{D}_S = \mathcal{D}_\pi$. *Thus, Markov* $\pi$*-strategies are sufficient in the problems (6) and (7).*

According to Theorems 1, 2, Markov $\pi$-strategies or Markov standard $\xi$-strategies are also sufficient in other (constrained) optimization problems where the objectives are expressed in terms of the occupation measures $\{\eta_n\}_{n=1}^\infty$; for example, in case of the following long-term average cost:

$$\varlimsup_{n\to\infty} \frac{\displaystyle\sum_{k=1}^n \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\eta_k(dx,da)}{\displaystyle\sum_{k=1}^n \eta_k(\mathbf{X}\times\mathbf{A})} \to \inf_{\{\eta_n\}_{n=1}^\infty \in \mathcal{D}_S}.$$

Moreover, the cost rates $c_i$ can also depend on the transition number $n$ (see (6)). This remark also concerns theorems 3 and 5.

# 4 Mixtures of simple deterministic Markov strategies

As was mentioned, the distribution $p_0$ is responsible for the mixtures. Suppose, for example, $S^1$ and $S^2$ are two simple deterministic Markov strategies defined by $\hat{\varphi}^1_n(x)$ and $\hat{\varphi}^2_n(x)$, $n = 1, 2, \ldots$ correspondingly, which give rise to the strategic measures $P^{S^1}_\gamma$ and$P^{S^2}_\gamma$ on the space

$$\Omega = (\mathbf{X}_\Delta \times \bar{\mathbb{R}}_+)^\infty \tag{10}$$

(see Remark 1 and the table at the end of Section 2). Now, take $\mathbf{\Xi} = \{1, 2\}$, $p_0(1) = p \geq 0$, $p_0(2) = 1 - p \geq 0$ and consider the $\xi$-strategy $S = \{\mathbf{\Xi}, p_0,\ \varphi_n(\xi_0, x) = \hat{\varphi}^{\xi_0}_n(x),\ n = 1, 2, \ldots\}$. (Components $p_n$ are of no importance here.) This will be an elementary mixture of two simple deterministic Markov strategies.

In the proof of Theorem 3, we construct the most general mixture of simple deterministic Markov strategies (see also Definition 5).

**Theorem 3** *Let*

$$\begin{aligned}\mathcal{D}_{dm} \quad = \quad & \{\ \{\eta^S_n\}_{n=1}^\infty,\ S = \{\mathbf{\Xi}^0 \times \mathbf{A}, \hat{p}_0(d\xi^0_0), \hat{p}_n(da_n|\xi^0_0, x_{n-1}),\ n = 1, 2, \ldots\} \\ & \textit{are mixtures of simple deterministic Markov strategies } \{\hat{\varphi}^{\xi^0_0}_n,\ n = 1, 2, \ldots\}\ \}.\end{aligned}$$

*and*

$$\begin{aligned}\mathcal{D}_{st} \quad = \quad & \{\ \{\eta^S_n\}_{n=1}^\infty,\ S = \{\mathbf{\Xi}^0 \times \mathbf{A}, \hat{p}_0(d\xi^0_0),\ \hat{p}_n(da_n|h_{n-1}),\ n = 1, 2, \ldots\} \\ & \textit{are mixtures of standard } \xi\textit{-strategies}\}.\end{aligned}$$

*Then* $\mathcal{D}_\xi = \mathcal{D}_{dm} = \mathcal{D}_{st}$.

# 5 Non-conservative transition rate and discounting

The possible gap

$$\alpha(x, a) \stackrel{\triangle}{=} q_x(a) - q(\mathbf{X} \setminus \{x\}|x, a) = q(\{\Delta\}|x, a) \geq 0$$

can be understood as the discount factor.

Let us denote $\hat{q}_x(a) \stackrel{\triangle}{=} q(\mathbf{X} \setminus \{x\}|x, a)$ and, for an arbitrary $\pi$-$\xi$-strategy $S$, consider the jump intensities

$$\hat{\lambda}_n(\Gamma|h_{n-1}, \xi_n, u) \stackrel{\triangle}{=} \lambda_n(\Gamma \cap \mathbf{X}|h_{n-1}, \xi_n, u)$$

and

$$\begin{aligned}\hat{\Lambda}_n(\Gamma, h_{n-1}, \xi_n, t) \quad &= \quad \Lambda_n(\Gamma \cap \mathbf{X}, h_{n-1}, \xi_n, t) \\ &= \quad \Lambda_n(\Gamma, h_{n-1}, \xi_n, t) - \int_{(0,t]} \int_{\mathbf{A}} \alpha(x_{n-1}, a)\pi_n(da|h_{n-1}, \xi_n, u)\ du.\end{aligned}$$

For the same spaces $\Omega$ and $\mathbf{H}_n$, we construct the strategic measure $\hat{P}^S_\gamma$ (with the corresponding expectation $\hat{E}^S_\gamma$) using stochastic kernels $\hat{G}$ defined by the same formulae (5), where $\lambda$ and $\Lambda$ are replaced with $\hat{\lambda}$ and $\hat{\Lambda}$. The only difference with $P^S_\gamma$ is that now the artificial state $\Delta$ never appears together with a finite sojourn time $\theta$. In other words, the controlled process does not escape from the state space at a finite time moment.

**Theorem 4** *For any* $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$, $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A})$,

$$\eta^S_n(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = \hat{E}^S_\gamma \left[ \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}}|H_{n-1}, \Xi_n, t - T_{n-1}) e^{-B(t)} dt \right],$$

*where*

$$B(t) = I\{X(t) \in \mathbf{X}\} \int_{(0,t]} \int_{\mathbf{A}} \alpha(X(u), a)\pi(da|u)du$$

*is the (random) discounting process.*

Now formula (6) takes the form:

$$W_0(S) = \sum_{n=1}^{\infty} \int_{\mathbf{X}\times\mathbf{A}} c_0(x,a)\eta_n^S(dx,da) = \hat{E}_\gamma^S \left[ \int_{(0,T_\infty)} \int_{\mathbf{A}} \pi(da|t) c_0(X(t),a) e^{-B(t)} dt \right] \to \inf_{S\in\Pi_S} .$$

In the simplest case $\alpha(x,a) \equiv \alpha > 0$ we have the standard discounted model investigated e.g. in [7, 11, 20].

# 6 Sufficiency of $\xi$-strategies, general case

Example presented in Section 7 shows that, if Condition 1 is not satisfied, then it can happen that, for a $\pi$-strategy $S$, there is no equivalent Markov standard $\xi$-strategy having the same occupation measures. Below, we describe a more general class of $\xi$-strategies which turns to be sufficient in the general case.

**Definition 7** *A Poisson-related $\xi$-strategy $S = \{\Xi, \varepsilon, \tilde{p}_{n,k}(da|x_{n-1}),\ n = 1,2,\dots,\ k = 0,1,2,\dots\}$ is defined by a constant $\varepsilon > 0$ and a sequence of stochastic kernels $\tilde{p}_{n,k}(da|x)$ from $\mathbf{X}_\Delta$ to $\mathbf{A}_\Delta$ with $\tilde{p}_{n,k}(\mathbf{A}(x)|x) = 1$. Here $\Xi = (\mathbb{R}\times\mathbf{A})^\infty$, and for $n = 1,2,\dots$ the distribution $p_n$ of $\Xi_n = (\tau_0^n, \alpha_0^n, \tau_1^n, \alpha_1^n, \dots)$ given $\mathbf{H}_{n-1}$ is defined as follows:*

- $p_n(\tau_0^n = 0|h_{n-1}) = 1$*; for* $i \geq 1$, $p_n(\tau_i^n \leq t|h_{n-1}) = 1 - e^{-\varepsilon t}$*;*
- *for all* $k \geq 0$, $p_n(\alpha_k^n \in \Gamma_{\mathbf{A}}|h_{n-1}) = \tilde{p}_{n,k}(\Gamma_{\mathbf{A}}|x_{n-1})$*;*
- *finally,*

$$\varphi_n(\xi_0, x_0, \xi_1, \theta_1, \dots, x_{n-1}, \xi_n, t - T_{n-1}) = I\{\tau_0^n + \dots + \tau_k^n < t - T_{n-1} \leq \tau_0^n + \dots + \tau_{k+1}^n\}\alpha_k^n.$$

*The $\xi_0$ component plays no role and is omitted. Note, function $\varphi_n$ does not depend on $\xi_0, x_0, \dots, x_{n-1}$ and is denoted as $\varphi_n(\xi_n, t - T_{n-1})$ in the proof of Theorem 5.*

Such a strategy means that, after any jump of the controlled process $X(t)$, we simulate a Poisson process and apply different randomized controls during the different sojourn times of that Poisson process.

**Theorem 5** *For any control strategy $S$, there is a Poisson-related $\xi$-strategy $S^P$ such that $\{\eta_n^S\}_{n=1}^\infty = \{\eta_n^{S^P}\}_{n=1}^\infty$. The value of $\varepsilon > 0$ can be chosen arbitrarily.*

# 7 Examples

Example 1 shows that, if a $\pi$-strategy $S$ is stationary then the occupation measures $\{\eta_n^S\}_{n=1}^\infty$ may be not generated by a stationary standard $\xi$-strategy. The reverse statement is also correct: not any one sequence $\{\eta_n^{\tilde{S}}\}_{n=1}^\infty$, coming from a stationary standard $\xi$-strategy $\tilde{S}$, can be generated by a stationary $\pi$-strategy.

Let $\mathbf{X} = \{1\}$, $\mathbf{A} = \mathbf{A}(1) = \{a_1, a_2\}$, $\gamma(1) = 1$, $q_1(a_1) = \lambda > 0$, $q_1(a_2) = 0$. For an arbitrary stationary $\pi$-strategy $S$ we have,

$$\begin{aligned} \text{either } \eta_1^S(1,a_1) < \infty \quad &\text{and} \quad \eta_1^S(1,a_2) < \infty \ (\text{if } \pi(a_1|1) > 0\ ), \\ \text{or } \eta_1^S(1,a_1) = 0 \quad &\text{and} \quad \eta_1^S(1,a_2) = \infty \ (\text{if } \pi(a_1|1) = 0\ ). \end{aligned}$$

If, for a stationary standard $\xi$-strategy $\tilde{S}$, $p(a_2|1) \in (0,1)$ then $\eta_1^{\tilde{S}}(1,a_1) = \frac{1-p(a_2|1)}{\lambda} \in (0,\infty)$, $\eta_1^{\tilde{S}}(1,a_2) = \infty$ and $\eta_1^{\tilde{S}}$ cannot be generated by a stationary $\pi$-strategy. If $\pi(a_1|1) \in (0,1)$ then $\eta_1^S(1,a_1) \in (0,\infty)$, $\eta_1^S(1,a_2) \in (0,\infty)$ and such an occupation measure cannot be generated by a stationary standard $\xi$-strategy.

Example 2 illustrates that Markov standard $\xi$-strategies (as well as stationary standard $\xi$-strategies and stationary $\pi$-strategies) are not sufficient in optimization problems.

Consider the following continuous-time Markov decision process, very similar to the one described in [9, Ex.3.1]. $\mathbf{X} = \{1\}$, $\mathbf{A} = \mathbf{A}(1) = (0, 1]$, $\gamma(1) = 1$, $q_1(a) = a$, $c_0(x, a) = a$, $N = 0$. Note that $q(\mathbf{X} \setminus \{1\}|1, a) = 0$ and $q(\mathbf{X}|1, a) = -q_1(a) = -a < 0$. After introducing the cemetery $\Delta$ with $\alpha(1, a) = q(\{\Delta\}|1, a) = q_1(a)$, we obtain the standard conservative transition rate $q$. In this model, we have a single sojourn time $\Theta = T$, so that the $n$ index is omitted.

It is obvious that, for any Markov standard $\xi$-strategy $p^M$ (which is also stationary),

$$\eta^{p^M}(\{1\} \times \Gamma_{\mathbf{A}}) = E_\gamma^{p^M}\left[\int_{(0,T]\cap\mathbb{R}_+} I\{A(t) \in \Gamma_{\mathbf{A}}\}dt\right] = \int_{\Gamma_{\mathbf{A}}} p^M(da|1) \cdot \frac{1}{a}$$

and

$$W_0(p^M) = E_\gamma^{p^M}\left[\int_{(0,T]\cap\mathbb{R}_+} A(t)dt\right] = \int_{\mathbf{A}} a\ \eta^{p^M}(\{1\} \times da) = \int_{\mathbf{A}} a\frac{1}{a}p^M(da|1) = 1.$$

For an arbitrary stationary $\pi$-strategy $S_\pi$, we similarly obtain

$$\eta^{S_\pi}(\{1\} \times \Gamma_{\mathbf{A}}) = \pi(\Gamma_{\mathbf{A}}) \Big/ \int_{\mathbf{A}} a\ \pi(da)$$

and

$$W_0(S_\pi) = \int_{\mathbf{A}} a\ \eta^{S_\pi}(\{1\} \times da) = 1.$$

On the other hand, under an arbitrarily fixed $\kappa > 0$, for the purely deterministic strategy $\varphi(1, u) = e^{-\kappa u}$, the (first) sojourn time $\Theta = T$ has the cumulative distribution function (CDF) $1 - e^{\frac{-1+e^{-\kappa\theta}}{\kappa}}$, so that $P_\gamma^\varphi(\Theta = \infty) = e^{-\frac{1}{\kappa}}$. Under an arbitrarily fixed $U \in (0, 1]$ we have

$$\begin{aligned}
\eta^\varphi(\{1\} \times (U, 1]) &= E_\gamma^\varphi\left[\int_{(0,\Theta]\cap\mathbb{R}_+} I\{e^{-\kappa u} \in (U, 1]\}du\right] = E_\gamma^\varphi\left[\int_{[e^{-\kappa\Theta},1)\cap\mathbb{R}_+} I\{y \in (U, 1]\}dy/(\kappa y)\right] \\
&= \frac{1}{\kappa}\int_{-\frac{\ln U}{\kappa}}^{\infty}[-\ln U](e^{-\kappa\theta} \cdot e^{\frac{-1+e^{-\kappa\theta}}{\kappa}})d\theta + \frac{1}{\kappa}\int_0^{-\frac{\ln U}{\kappa}} \kappa\theta(e^{-\kappa\theta} \cdot e^{\frac{-1+e^{-\kappa\theta}}{\kappa}})d\theta \\
&\quad + \frac{1}{\kappa}[-\ln U] \cdot e^{-\frac{1}{\kappa}} = [-\ln U]\frac{1}{\kappa}(-e^{-\frac{1}{\kappa}} + e^{\frac{U-1}{\kappa}}) + \theta\left[1 - e^{\frac{-1+e^{-\kappa\theta}}{\kappa}}\right]\Big|_0^{-\frac{\ln U}{\kappa}} \\
&\quad - \int_0^{-\frac{\ln U}{\kappa}}\left[1 - e^{\frac{-1+e^{-\kappa\theta}}{\kappa}}\right]d\theta + \frac{1}{\kappa}[-\ln U] \cdot e^{-\frac{1}{\kappa}} = \int_0^{-\frac{\ln U}{\kappa}} e^{\frac{-1+e^{-\kappa\theta}}{\kappa}}d\theta \\
&= \int_U^1 \frac{e^{\frac{-1+a}{\kappa}}}{\kappa a}da,
\end{aligned}$$

so that measure $\eta^\varphi(\{1\} \times da)$ is absolutely continuous w.r.t. the Lebesgue measure, the density being $\frac{e^{\frac{-1+a}{\kappa}}}{\kappa a}$ and

$$W_0(\varphi) = \int_{\mathbf{A}} a\ \eta^\varphi(\{1\} \times da) = 1 - e^{-\frac{1}{\kappa}}. \tag{11}$$

According to Theorem 1, there is a Markov standard $\xi$-strategy $S_\xi$ such that $\eta^{S_\xi} \geq \eta^\varphi$. It is given by formula (16). One can also build the Poisson-related $\xi$-strategy $S^P$ such that $\eta^{S^P} = \eta^\varphi$, using the proof of Theorem 5. The detailed calculations can be found in [22]. Finally, it is clear that $\inf_{S\in\Pi_S} W_0(S) = 0$: see (11) with $\kappa \to \infty$, but the optimal strategy does not exist because $\Theta > 0$ and $c_0(x, a) > 0$. Note also that, if we extend the action space to $[0, 1]$ and keep $q_1$ and $c_0$ continuous, i.e., $q_1(0) = c_0(0) = 0$, then stationary deterministic strategy $\varphi^*(x) = 0$ is optimal with $W_0(\varphi^*) = 0$.

# 8 Acknowledgement

The author is thankful to Prof. F.Dufour and Dr. Y.Zhang for fruitful discussions and careful reading of the draft of this article.

# 9 Conclusion

In the optimal control theory, the researchers traditionally start with a wide class of control strategies and prove the sufficiency of a small collection of easily implementable strategies, e.g., a unique strategy, if a particular problem is exactly solved. For example, in [10, 11, 20, 23, 25], starting from general relaxed strategies, the authors prove the sufficiency of stationary deterministic strategies (stationary relaxed strategies in constrained problems). In the current article, the new very general set of control strategies is introduced, and a series of theorems state the sufficiency of Markov relaxed, randomized, Poisson-related strategies and mixtures of Markov deterministic strategies. Note, the cost rate and the transition rate can be unbounded and accumulation of jumps is not excluded.

Theorem 5 about sufficiency of Poisson-related strategies can be a starting point for involving the results in discrete-time Markov decision processes (DTMDP) like the Linear Programming approach developed e.g. in [13, 18]. Under very mild conditions, it will be possible to prove the sufficiency of stationary randomized strategies. Remember, Example 2 in Section 7 shows that, in general, stationary strategies are not sufficient in optimization problems. This fact is known also in the discrete-time case [19, §§2.2.11,2.2.12,2.2.13]. Transformation to discrete time is a well known trick [21]. In this connection, Theorem 5 will lead to the DTMDP with possible transitions to the same state (loops). These ideas will be developed in [22].

We consider the sufficiency of randomized and Poisson-related strategies more valuable compared with the traditional relaxed strategies because the latter ones cannot be realized on practice if they are not purely deterministic: the trajectories of the action process are not measurable. The word "sufficient" refers to the total expected cost/reward. If one is also interested in the variance of the total cost, then the current results and conclusions are not relevant.

# 10 Appendix

Proof of Lemma 1. For any fixed $u$, the mapping $\xi_n \to \xi_n(u)$ is measurable [5, Ch.3,Prop.7.1], so that $\xi_n(u)$ is a right continuous random process defined on $D_{\mathbf{A}}[0,\infty)$. It is progressively measurable, e.g. if we consider the trivial filtration $\mathcal{G}_u \equiv \mathcal{B}(D_{\mathbf{A}}[0,\infty))$ [3, T11]; hence the mapping $(\xi_n, u) \to \xi_n(u)$ is $\mathcal{B}(D_{\mathbf{A}}[0,\infty)\times\mathbb{R}_+)$-measurable. ■

Proof of Theorem 1. Inclusion $\mathcal{D}_\xi \subset \mathcal{D}_S$ is obvious.

Let us prove that $\mathcal{D}_S \subset \mathcal{D}_\xi$ if Condition 1-(b) is satisfied. Simultaneously, we will establish the first assertion of the theorem assuming that $q_x(a) > 0$ for all $(x,a) \in \mathbb{K}$.

Let $S = \{\mathbf{\Xi}, p_0, \langle p_n, \pi_n\rangle,\ n = 1,2,\ldots\}$ be an arbitrary $\pi$-$\xi$-strategy and introduce the following *occupancy* measures ($n = 1,2,\ldots$) on $\mathbf{X}\times\mathbf{A}$

$$\rho_n^S(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}}) = E_\gamma^S\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+} I\{X_{n-1}\in\Gamma_{\mathbf{X}}\}\int_{\Gamma_{\mathbf{A}}}\pi_n(da|H_{n-1},\Xi_n,t-T_{n-1})q_{X_{n-1}}(a)dt\right],$$

$\Gamma_{\mathbf{X}}\in\mathcal{B}(\mathbf{X}), \Gamma_{\mathbf{A}}\in\mathcal{B}(\mathbf{A})$. Note that $\rho_n^S(\mathbf{X}\times\mathbf{A}) = 0$ if and only if $X_{n-1} = \Delta$ $P_\gamma^S$-a.s.

First of all, let us show that these measures are finite for all $n = 1,2,\ldots$ even if the jump intensity $q_x(a)$ is unbounded. Let $\pi_u(\cdot) \stackrel{\triangle}{=} \pi_n(\cdot|h_{n-1},\xi,u) \in \mathcal{P}(\mathbf{A})$ assuming $x_{n-1}\neq\Delta$, and

introduce the following finite measures (depending on $h_{n-1}, \xi$) on $\mathcal{P}(\mathbf{A})$:

$$\begin{aligned} k_n(\Gamma_\pi, h_{n-1}, \xi) &= I\{x_{n-1} \neq \Delta\} \int_{(0,\infty)} I\{\pi_\theta \in \Gamma_\pi\} \tilde{G}_n(d\theta \times \mathbf{X}_\Delta | h_{n-1}, \xi), \\ K_n(\Gamma_\pi, h_{n-1}, \xi) &= I\{x_{n-1} \neq \Delta\} \int_{(0,\infty)} \int_{(0,\theta]} I\{\pi_u \in \Gamma_\pi\} du \cdot \tilde{G}_n(d\theta \times \mathbf{X}_\Delta | h_{n-1}, \xi), \\ & \qquad \Gamma_\pi \in \mathcal{B}(\mathcal{P}(\mathbf{A})), \quad n = 1, 2, \ldots \end{aligned}$$

Here

$$\begin{aligned} \tilde{G}_n(\{\infty\} \times \mathbf{X}_\Delta | h_{n-1}, \xi) &= \delta_{x_{n-1}}(\mathbf{X}) e^{-\Lambda_n(\mathbf{X}_\Delta, h_{n-1}, \xi, \infty)}; \\ \tilde{G}_n(\Gamma_\mathbb{R} \times \mathbf{X}_\Delta | h_{n-1}, \xi) &= \delta_{x_{n-1}}(\mathbf{X}) \int_{\Gamma_\mathbb{R}} \lambda_n(\mathbf{X}_\Delta | h_{n-1}, \xi, t) \\ & \quad \times e^{-\Lambda_n(\mathbf{X}_\Delta, h_{n-1}, \xi, t)} dt \quad \text{for } \Gamma_\mathbb{R} \in \mathcal{B}(\mathbb{R}_+). \end{aligned}$$

Then, according to Lemma 4.3 [7],

$$k_n(\Gamma_\pi, h_{n-1}, \xi) = \int_{\Gamma_\pi} \left[ \int_{\mathbf{A}} q_{x_{n-1}}(a) \pi(da) \right] K_n(d\pi, h_{n-1}, \xi). \tag{12}$$

(Here $\pi \in \mathcal{P}(\mathbf{A})$ and $\int_{\mathbf{A}} q_{x_{n-1}}(a)\pi(da)$ play the role of $a$ and $q(a)$ in [7] correspondingly.) Now, since function $q_{x_{n-1}}(a)$ is non-negative, according to (12), we have

$$\begin{aligned} & \int_{\mathcal{P}(\mathbf{A})} \left[ \int_{\mathbf{A}} q_{x_{n-1}}(a) \pi(da) \right] K_n(d\pi, h_{n-1}, \xi) \\ = & \int_{(0,\infty)} I\{x_{n-1} \neq \Delta\} \left[ \int_{(0,\theta]} \int_{\mathcal{P}(\mathbf{A})} \delta_{\pi_s}(d\pi) \left[ \int_{\mathbf{A}} q_{x_{n-1}}(a) \pi(da) \right] ds \right] \tilde{G}_n(d\theta \times \mathbf{X}_\Delta | h_{n-1}, \xi) \\ = & \; k_n(\mathcal{P}(\mathbf{A}) | h_{n-1}, \xi) = \tilde{G}_n(\mathbb{R}_+ \times \mathbf{X}_\Delta | h_{n-1}, \xi) \leq 1, \end{aligned} \tag{13}$$

so that

$$\begin{aligned} \rho_n^S(\mathbf{X} \times \mathbf{A}) &= E_\gamma^S \left[ \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} I\{X_{n-1} \neq \Delta\} \int_{\mathbf{A}} \pi_n(da | H_{n-1}, \Xi_n, t - T_{n-1}) q_{X_{n-1}}(a) dt \right] \\ &= E_\gamma^S \left[ E_\gamma^S \left[ I\{X_{n-1} \neq \Delta\} \int_{(0, \Theta_n] \cap \mathbb{R}_+} \int_{\mathbf{A}} q_{X_{n-1}}(a) \pi_s(da) ds | H_{n-1}, \Xi_n \right] \right] \\ &= E_\gamma^S \left[ I\{X_{n-1} \neq \Delta\} k_n(\mathcal{P}(\mathbf{A}), H_{n-1}, \Xi_n) \right] \\ &= E_\gamma^S \left[ I\{X_{n-1} \neq \Delta\} \tilde{G}_n(\mathbb{R}_+ \times \mathbf{X}_\Delta | H_{n-1}, \Xi_n) \right] \leq 1, \end{aligned} \tag{14}$$

and the $\rho_n^S$ measure is finite. Remember, $\tilde{G}_n(\mathbb{R}_+ \times \mathbf{X}_\Delta | H_{n-1}, \Xi_n) > 0$ $P_\gamma^S$-a.s. if $X_{n-1} \neq \Delta$, because of Condition 1-(a).

For the measures

$$\begin{aligned} \hat{k}_n(\Gamma_\pi \times \Gamma_\mathbf{X}, h_{n-1}, \xi) &= I\{x_{n-1} \in \Gamma_\mathbf{X}\} \int_{(0,\infty)} I\{\pi_\theta \in \Gamma_\pi\} \tilde{G}_n(d\theta \times \mathbf{X}_\Delta | h_{n-1}, \xi), \\ \hat{K}_n(\Gamma_\pi \times \Gamma_\mathbf{X}, h_{n-1}, \xi) &= I\{x_{n-1} \in \Gamma_\mathbf{X}\} \int_{(0,\infty)} \int_{(0,\theta]} I\{\pi_u \in \Gamma_\pi\} du \cdot \tilde{G}_n(d\theta \times \mathbf{X}_\Delta | h_{n-1}, \xi) \end{aligned}$$

on $\mathcal{P}(\mathbf{A}) \times \mathbf{X}$, the similar calculations result in expressions

$$\begin{aligned} \int_{\mathcal{P}(\mathbf{A}) \times \Gamma_\mathbf{X}} \int_{\mathbf{A}} q_x(da) \pi(da) \hat{K}_n(d\pi, dx, h_{n-1}, \xi) &= \hat{k}_n(\mathcal{P}(\mathbf{A}) \times \Gamma_\mathbf{X}, h_{n-1}, \xi); \\ \rho_n^S(\Gamma_\mathbf{X} \times \mathbf{A}) = E_\gamma^S \left[ \hat{k}_n(\mathcal{P}(\mathbf{A}) \times \Gamma_\mathbf{X}, H_{n-1}, \Xi_n) \right] &= E_\gamma^S [I\{X_{n-1} \in \Gamma_\mathbf{X}\} \tilde{G}_n(\mathbb{R}_+ \times \mathbf{X}_\Delta | H_{n-1}, \Xi_n)], \\ & \qquad n = 1, 2, \ldots \end{aligned} \tag{15}$$

Having the occupancy measures $\rho_n^S$ in hand, we introduce the stochastic kernels $p_n^M$ (defined $\rho_n^S(\cdot,\mathbf{A})$-a.s.) coming from formula

$$\rho_n^S(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})=\int_{\Gamma_{\mathbf{X}}}\rho_n^S(dx\times\mathbf{A})p_n^M(\Gamma_{\mathbf{A}}|x).$$

Note that $\rho_n^S(\mathbf{X}\times\mathbf{A})=0$ if and only if $X_{n-1}=\Delta$ $P_\gamma^S$-a.s., and we put $p_n^M(\{\Delta\}|\Delta)=1$ as usual. For $x_{n-1}\neq\Delta$, one can provide the explicit formula for $p_n^M$:

$$p_n^M(\Gamma_{\mathbf{A}}|x_{n-1})=\frac{E_\gamma^S\left[\int_{(0,\Theta_n]\cap\mathbb{R}_+}\int_{\Gamma_{\mathbf{A}}}\pi_n(da|H_{n-1},\Xi_n,u)q_{X_{n-1}}(a)du|X_{n-1}=x_{n-1}\right]}{E_\gamma^S\left[\int_{(0,\Theta_n]\cap\mathbb{R}_+}\int_{\mathbf{A}}\pi_n(da|H_{n-1},\Xi_n,u)q_{X_{n-1}}(a)du|X_{n-1}=x_{n-1}\right]}. \tag{16}$$

Note, the denominator equals 1 under Condition 1(b). Equation (16) holds $\hat\rho_n^S$-a.s., where $\hat\rho_n^S(\Gamma_{\mathbf{X}})=\rho_n^S(\Gamma_{\mathbf{X}}\times\mathbf{A})$ is the marginal of $\rho_n^S$. Below we omit such remarks for equations involving conditional expectations.

Consider the Markov standard $\xi$-strategy $S_\xi=\{\mathbf{A},\ p_n^M,\ n=1,2,\ldots\}$. Let

$$\tilde\rho_n(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})\stackrel{\triangle}{=}E_\gamma^{S_\xi}\left[I\{X_{n-1}\in\Gamma_{\mathbf{X}},A_n\in\Gamma_{\mathbf{A}}\}\right]$$

be a measure on $\mathbf{X}\times\mathbf{A}$ and prove by induction that $\tilde\rho_n\geq\rho_n^S$. Equality $\tilde\rho_1(\Gamma_{\mathbf{X}}\times\mathbf{A})=\rho_1^S(\Gamma_{\mathbf{X}}\times\mathbf{A})=\gamma(\Gamma_{\mathbf{X}})$ is obvious. Assume $\tilde\rho_n(\Gamma_{\mathbf{X}}\times\mathbf{A})\geq\rho_n^S(\Gamma_{\mathbf{X}}\times\mathbf{A})$ for some $n\geq1$. Then, by the definition of the $\xi$-strategy $S_\xi$,

$$\tilde\rho_n(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})=\int_{\Gamma_{\mathbf{X}}}\tilde\rho_n(dx\times\mathbf{A})p_n^M(\Gamma_{\mathbf{A}}|x),$$

so that $\tilde\rho_n(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})\geq\rho_n^S(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})$ and it remains to show that $\tilde\rho_{n+1}(\Gamma_{\mathbf{X}}\times\mathbf{A})\geq\rho_{n+1}^S(\Gamma_{\mathbf{X}}\times\mathbf{A})$.

$$\begin{aligned}
\tilde\rho_{n+1}(\Gamma_{\mathbf{X}}\times\mathbf{A}) &= \int_{\mathbf{X}\times\mathbf{A}}\left[\int_{(0,\infty)}q(\Gamma_{\mathbf{X}}\setminus\{x\}|x,a)e^{-q_x(a)t}dt\right]\tilde\rho_n(dx,da)\\
&= \int_{\mathbf{X}\times\mathbf{A}}\frac{q(\Gamma_{\mathbf{X}}\setminus\{x\}|x,a)}{q_x(a)}\tilde\rho_n(dx,da)\\
&\geq E_\gamma^S\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}\int_{\mathbf{A}}q(\Gamma_{\mathbf{X}}\setminus\{X_{n-1}\}|X_{n-1},a)\pi_n(da|H_{n-1},\Xi_n,t-T_{n-1})dt\right]
\end{aligned}$$

because $\tilde\rho_n\geq\rho_n^S$. The cases $\tilde\rho_n=0$ or $\tilde\rho_{n+1}=0$ are not excluded.

On the other hand, using (12), we obtain

$$\begin{aligned}
&E_\gamma^S\left[I\{X_n\in\Gamma_{\mathbf{X}}\}|H_{n-1},\Xi_n\right]\\
= &\int_{\mathcal{P}(\mathbf{A})}\frac{\int_{\mathbf{A}}\pi(da)q(\Gamma_{\mathbf{X}}\setminus\{X_{n-1}\}|X_{n-1},a)}{\int_{\mathbf{A}}\pi(da)q_{X_{n-1}}(a)}k_n(d\pi,H_{n-1},\Xi_n)\\
= &\int_{\mathcal{P}(\mathbf{A})}\frac{\int_{\mathbf{A}}\pi(da)q(\Gamma_{\mathbf{X}}\setminus\{X_{n-1}\}|X_{n-1},a)}{\int_{\mathbf{A}}\pi(da)q_{X_{n-1}}(a)}\int_{\mathbf{A}}\pi(da)q_{X_{n-1}}(a)K_n(d\pi,H_{n-1},\Xi_n)\\
= &E_\gamma^S\left[\int_{(0,\Theta_n]\cap\mathbb{R}_+}\left[\int_{\mathbf{A}}\pi_s(da)q(\Gamma_{\mathbf{X}}\setminus\{X_{n-1}\}|X_{n-1},a)\right]ds|H_{n-1},\Xi_n\right],
\end{aligned}$$

so that, from (15) we have

$$\begin{aligned}
\rho_{n+1}^S(\Gamma_{\mathbf{X}}\times\mathbf{A}) &= E_\gamma^S\left[I\{X_n\in\Gamma_{\mathbf{X}}\}\tilde G_{n+1}(\mathbb{R}_+\times\mathbf{X}_\Delta|H_n,\Xi_{n+1})\right]\leq E_\gamma^S\left[I\{X_n\in\Gamma_{\mathbf{X}}\}\right]\\
&= E_\gamma^S\left[E_\gamma^S\left[\int_{(0,\Theta_n]\cap\mathbb{R}_+}\left[\int_{\mathbf{A}}\pi_s(da)q(\Gamma_{\mathbf{X}}\setminus\{X_{n-1}\}|X_{n-1},a)\right]ds|H_{n-1},\Xi_n\right]\right]\\
&= E_\gamma^S\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}\int_{\mathbf{A}}q(\Gamma_{\mathbf{X}}\setminus\{X_{n-1}\}|X_{n-1},a)\pi_n(da|H_{n-1},\Xi_n,t-T_{n-1})dt\right]\\
&\leq \tilde\rho_{n+1}(\Gamma_{\mathbf{X}}\times\mathbf{A}).
\end{aligned}$$

As a result, $\tilde{\rho}_n \geq \rho_n^S$ for all $n = 1, 2, \ldots$. All inequalities become equalities under Condition 1-(b) because here $\tilde{G}_n(\mathbb{R}_+ \times \mathbf{X}_\Delta | h_{n-1}, \xi) \equiv 1$.

Clearly, $\eta_n^S(\Gamma_\mathbf{X} \times \Gamma_\mathbf{A}) = \int_{\Gamma_\mathbf{X} \times \Gamma_\mathbf{A}} \left[\frac{1}{q_x(a)}\right] \rho_n^S(dx, da)$ and, to complete this part of the proof, it remains to notice that

$$\begin{aligned} \eta_n^{S_\xi}(\Gamma_\mathbf{X} \times \Gamma_\mathbf{A}) &= E_\gamma^{S_\xi}\left[ I\{X_{n-1} \in \Gamma_\mathbf{X}\} I\{A_n \in \Gamma_\mathbf{A}\} E_\gamma^{S_\xi}\left[ \int_0^{\Theta_n} ds | H_{n-1}, A_n \right]\right] \\ &= \int_{\Gamma_\mathbf{X} \times \Gamma_\mathbf{A}} \left[\frac{1}{q_x(a)}\right] \tilde{\rho}_n(dx, da). \end{aligned}$$

We have proved that $\eta_n^S \leq \eta_n^{S_\xi}$ for all $n = 1, 2, \ldots$. Under Condition 1-(b), we have equality, so that $\mathcal{D}_S = \mathcal{D}_\xi$. ■

For the proof of Corollary 1, it is sufficient to notice that, for quasi-stationary strategy $S$, expression (16) for $p_n^M$ does not depend on $n$. ■

Proof of Theorem 2. For an arbitrarily fixed $\pi$-$\xi$-strategy $S = \{\mathbf{\Xi}, p_0, \langle p_n, \pi_n \rangle,\ n = 1, 2, \ldots\}$, introduce the following purely relaxed Markov strategy $\tilde{S} = \{\pi_n^M,\ n = 1, 2, \ldots\}$:

$$\pi_n^M(\Gamma_\mathbf{A} | x_{n-1}, s) = \frac{E_\gamma^S[\pi_n(\Gamma_\mathbf{A} | H_{n-1}, \Xi_n, s) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, s)} | X_{n-1} = x_{n-1}]}{E_\gamma^S[e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, s)} | X_{n-1} = x_{n-1}]}. \tag{17}$$

Firstly, let us prove that, for any $n = 0, 1, \ldots$, the following joint distributions coincide

$$E_\gamma^S[I\{\Theta_n \in \Gamma_\mathbb{R}\} I\{X_n \in \Gamma_\mathbf{X}\}] = E_\gamma^{\tilde{S}}[I\{\Theta_n \in \Gamma_\mathbb{R}\} I\{X_n \in \Gamma_\mathbf{X}\}], \tag{18}$$

$\Gamma_\mathbb{R} \in \mathcal{B}(\bar{\mathbb{R}}_+)$, $\Gamma_\mathbf{X} \in \mathcal{B}(\mathbf{X}_\Delta)$.

Formula (18) is valid for $n = 0$. (We always put $\Theta_0 \equiv 0$.) Suppose it holds for some $n - 1 \geq 0$. Below, $\lambda_n^M$ and $\Lambda_n^M$ correspond to $\pi_n^M$; these functions, except for $\Gamma \in \mathcal{B}(\mathbf{X}_\Delta)$, depend only on $x_{n-1}$ and $s$ (or $t$). Since

$$\int_{(0,t] \cap \mathbb{R}_+} \lambda_n(\mathbf{X}_\Delta | h_{n-1}, \xi_n, s) e^{-\Lambda_n(\mathbf{X}_\Delta, h_{n-1}, \xi_n, s)} ds = 1 - e^{-\Lambda_n(\mathbf{X}_\Delta, h_{n-1}, \xi_n, t)}$$

and according to the Fubini Theorem, we have

$$\begin{aligned} & E_\gamma^S\left[ e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1} \right] \\ = \ & 1 - \int_{(0,t]} E_\gamma^S\left[ \lambda_n(\mathbf{X}_\Delta | H_{n-1}, \Xi_n, s) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, s)} | X_{n-1} = x_{n-1} \right] ds. \end{aligned}$$

This and other equalities below hold for $E_\gamma^S$-almost all $x_{n-1}$ and for $E_\gamma^{\tilde{S}}$-almost all $x_{n-1}$. Therefore, the derivative $\frac{d}{dt} \ln\left(E_\gamma^S\left[e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1}\right]\right)$ is well defined for almost all $t$ and equals

$$\frac{-E_\gamma^S\left[\lambda_n(\mathbf{X}_\Delta | H_{n-1}, \Xi_n, t) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1}\right]}{E_\gamma^S[e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1}]} = -\lambda_n^M(\mathbf{X}_\Delta | X_{n-1} = x_{n-1}, t),$$

so that

$$\Lambda_n^M(\mathbf{X}_\Delta, x_{n-1}, t) = -\ln\left(E_\gamma^S\left[e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1}\right]\right)$$

and

$$e^{-\Lambda_n^M(\mathbf{X}_\Delta, x_{n-1}, t)} = E_\gamma^S\left[e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1}\right]. \tag{19}$$

Now, for any $\Gamma_\mathbf{X} \in \mathcal{B}(\mathbf{X}_\Delta)$,

$$\lambda_n^M(\Gamma_\mathbf{X} | x_{n-1}, t) e^{-\Lambda_n^M(\mathbf{X}_\Delta, x_{n-1}, t)} = E_\gamma^S\left[\lambda_n(\Gamma_\mathbf{X} | H_{n-1}, \Xi_n, t) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t)} | X_{n-1} = x_{n-1}\right]$$

due to the definition of the $\pi_n^M$ kernel. Therefore, the conditional distributions

$$E_\gamma^S[I\{\Theta_n \in \Gamma_{\mathbb{R}}\}I\{X_n \in \Gamma_{\mathbf{X}}\}|X_{n-1}=x_{n-1}] = E_\gamma^{\tilde{S}}[I\{\Theta_n \in \Gamma_{\mathbb{R}}\}I\{X_n \in \Gamma_{\mathbf{X}}\}|X_{n-1}=x_{n-1}]$$

coincide and formula (18) holds for $n$ by induction.

Since, by the Fubini Theorem,

$$\begin{aligned}
&\int_{(0,\infty)} \lambda_n(\mathbf{X}_\Delta|h_{n-1},\xi_n,\theta)e^{-\Lambda_n(\mathbf{X}_\Delta,h_{n-1},\xi_n,\theta)}\left[\int_{(0,\theta]}\pi_n(\Gamma_{\mathbf{A}}|h_{n-1},\xi_n,u)du\right]d\theta \\
= &\int_{(0,\infty)}\left[\int_{[u,\infty)}\lambda_n(\mathbf{X}_\Delta|h_{n-1},\xi_n,\theta)e^{-\Lambda_n(\mathbf{X}_\Delta,h_{n-1},\xi_n,\theta)}\pi_n(\Gamma_{\mathbf{A}}|h_{n-1},\xi_n,u)\ d\theta\right]du \qquad (20)\\
= &\int_{(0,\infty)} e^{-\Lambda_n(\mathbf{X}_\Delta,h_{n-1},\xi_n,u)}\pi_n(\Gamma_{\mathbf{A}}|h_{n-1},\xi_n,u)\ du,
\end{aligned}$$

we conclude that, for any $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$, $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A})$,

$$\begin{aligned}
&E_\gamma^S\left[I\{X_{n-1}\in\Gamma_{\mathbf{X}}\}\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}\pi_n(\Gamma_{\mathbf{A}}|H_{n-1},\Xi_n,t-T_{n-1})dt|X_{n-1}=x_{n-1}\right]\\
= &\ I\{x_{n-1}\in\Gamma_{\mathbf{X}}\}E_\gamma^S\left[\int_{(0,\infty)}e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,u)}\pi_n(\Gamma_{\mathbf{A}}|H_{n-1},\Xi_n,u)du|X_{n-1}=x_{n-1}\right]\\
= &\ I\{x_{n-1}\in\Gamma_{\mathbf{X}}\}\int_{(0,\infty)}\pi_n^M(\Gamma_{\mathbf{A}}|x_{n-1},u)\cdot E_\gamma^S\left[e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,u)}|X_{n-1}=x_{n-1}\right]\ du\\
= &\ I\{x_{n-1}\in\Gamma_{\mathbf{X}}\}\int_{(0,\infty)}\pi_n^M(\Gamma_{\mathbf{A}}|x_{n-1},u)e^{-\Lambda_n^M(\mathbf{X}_\Delta,x_{n-1},u)}du
\end{aligned}$$

(see (19) ), and the last expression, similarly to (20), equals

$$E_\gamma^{\tilde{S}}\left[I\{X_{n-1}\in\Gamma_{\mathbf{X}}\}\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}\pi_n^M(\Gamma_{\mathbf{A}}|X_{n-1},t-T_{n-1})dt|X_{n-1}=x_{n-1}\right].$$

Therefore,

$$\begin{aligned}
&\eta_n^S(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})\\
= &\int_{\mathbf{X}_\Delta}E_\gamma^S\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}I\{X_{n-1}\in\Gamma_{\mathbf{X}}\}\pi_n(\Gamma_{\mathbf{A}}|H_{n-1},\Xi_n,t-T_{n-1})dt|X_{n-1}=x_{n-1}\right]m(dx_{n-1})\\
= &\ \eta_n^{\tilde{S}}(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}}),
\end{aligned}$$

where $m(\Gamma)\stackrel{\triangle}{=}E_\gamma^S[I\{X_{n-1}\in\Gamma\}]=E_\gamma^{\tilde{S}}[I\{X_{n-1}\in\Gamma\}]$ (see (18)).

In case $T_{n-1}<\infty$ and $T_n=\infty$, the integration is over the open interval $(T_{n-1},\infty)$. ■

Proof of Theorem 3. Before starting the proof itself, we need several additional constructions.

For an arbitrary simple deterministic Markov strategy $S=\{\hat{\varphi}_n,\ n=1,2,\ldots\}$, let

$$\hat{\omega}(\omega)=(x_0,a_1=\hat{\varphi}_1(x_0),\theta_1,x_1,a_2=\hat{\varphi}_2(x_1),\theta_2,\ldots) \qquad (21)$$

be the mapping from $\Omega$ to

$$\hat{\Omega}\stackrel{\triangle}{=}(\mathbf{X}_\Delta\times\mathbf{A}_\Delta\times\bar{\mathbb{R}}_+)^\infty. \qquad (22)$$

Let $\hat{P}_\gamma^S$ be the image of $P_\gamma^S$ w.r.t. this mapping and $\hat{E}_\gamma^S$ be the expectation w.r.t. this probability measure. Note that, if $X_n=\Delta$, then $\hat{P}_\gamma^S$-a.s. $A_{n+1}=\Delta,\Theta_{n+1}=\infty,X_{n+1}=\Delta$.

Now

$$\eta_n^S(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})=\hat{E}_\gamma^S\left[\Theta_n\ I\{X_{n-1}\in\Gamma_{\mathbf{X}}\}I\{A_n\in\Gamma_{\mathbf{A}}\}\right]. \qquad (23)$$

The same formula is valid for a standard $\xi$-strategy $S = \{\mathbf{A},\ p_n^M,\ n = 1, 2, \ldots\}$. Here, one does not need to introduce the mapping $\hat{\omega}(\omega)$ because, for standard $\xi$-strategies, the sample space already has the form $\hat{\Omega}$. Nevertheless, we keep the notations $\hat{P}_\gamma^S = P_\gamma^S$ and $\hat{E}_\gamma^S = E_\gamma^S$ for the further convenience.

According to the definition of the strategic measures, if $S$ is a simple deterministic Markov strategy or a standard $\xi$-strategy, then for arbitrary $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X}_\Delta)$, $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A}_\Delta)$, $\Gamma_{\mathbb{R}} \in \mathcal{B}(\bar{\mathbb{R}}_+)$, we have $\hat{P}_\gamma^S(X_0 \in \Gamma_{\mathbf{X}}) = \gamma(\Gamma_{\mathbf{X}})$;

$$\hat{P}_\gamma^S(A_n \in \Gamma_{\mathbf{A}}|X_0, A_1, \Theta_1, \ldots, X_{n-1}) = p_n^M(\Gamma_{\mathbf{A}}|H_{n-1}) \tag{24}$$

$(= I\{\Gamma_{\mathbf{A}} \ni \hat{\varphi}_n(X_{n-1})\}$ in case the strategy $S$ is simple deterministic Markov$)$

$$\hat{P}_\gamma^S(\Theta_n \in \Gamma_{\mathbb{R}}|X_0, A_1, \Theta_1, \ldots, X_{n-1}, A_n) = I\{X_{n-1} \neq \Delta\} \int_{\Gamma_{\mathbb{R}} \cap \mathbb{R}_+} q_{X_{n-1}}(A_n) e^{-q_{X_{n-1}}(A_n)t} dt$$

$$+ I\{q_{X_{n-1}}(A_n) = 0 \ or \ X_{n-1} = \Delta\} I\{\Gamma_{\mathbb{R}} \ni \infty\}, \tag{25}$$

$$\hat{P}_\gamma^S(X_n \in \Gamma_{\mathbf{X}}|X_0, A_1, \Theta_1, \ldots, X_{n-1}, A_n, \Theta_n), \tag{26}$$

$$= I\{X_{n-1} \neq \Delta\} \frac{q(\Gamma_{\mathbf{X}} \setminus \{X_{n-1}\}|X_{n-1}, A_n)}{q_{X_{n-1}}(A_n)} + I\{q_{X_{n-1}}(A_n) = 0 \ or \ X_{n-1} = \Delta\} I\{\Gamma_{\mathbf{X}} \ni \Delta\}$$

where $\dfrac{0}{0} \triangleq 0$.

Formulae (24) and (26) define the marginal of the measure $\hat{P}_\gamma^S$ on $(\mathbf{X}_\Delta \times \mathbf{A}_\Delta)^\infty$ denoted below as $\hat{P}_\gamma^{S\mathcal{M}}$, and formula (25) makes it possible to reconstruct $\hat{P}_\gamma^S$ having $\hat{P}_\gamma^{S\mathcal{M}}$.

Let us show that $\mathcal{D}_{st} \subset \mathcal{D}_\xi$. For a fixed mixture $S = \{\mathbf{\Xi}^0 \times \mathbf{A}, \hat{p}_0(d\xi_0^0), \hat{p}_n(da_n|\xi_0^0, x_{n-1}),\ n = 1, 2, \ldots\}$ of standard $\xi$-strategies, we define

$$\hat{P}_\gamma^S(d\hat{\omega}) = P_\gamma^S(\mathbf{\Xi}^0 \times d\hat{\omega}) = \int_{\mathbf{\Xi}^0} \hat{p}_0(d\xi_0^0) \hat{P}_\gamma^{S^{\xi_0^0}}(d\hat{\omega}),$$

where $S^{\xi_0^0} = \{\mathbf{A}, \hat{p}_n(da_n|\xi_0^0, x_0, a_1, \theta_1, \ldots, x_{n-1})\ n = 1, 2, \ldots\}$ is a specific Markov standard $\xi$-strategy under a fixed $\xi_0^0 \in \mathbf{\Xi}^0$. Note that the $\hat{P}_\gamma^{\xi_0^0}$ measure is measurable w.r.t. $\xi_0^0$ [12, C.10]. Recall that, according to Remark 1, the measure $P_\gamma^S$ is defined on $\mathbf{\Xi}^0 \times \hat{\Omega}$ and the measures $\hat{P}_\gamma^{S^{\xi_0^0}} = P_\gamma^{S^{\xi_0^0}}$ are defined on $\hat{\Omega}$: see (22) and the table at the end of Section 2. Like previously, $\hat{P}_\gamma^{S\mathcal{M}}$ is the marginal of $\hat{P}_\gamma^S$ on $(\mathbf{X}_\Delta \times \mathbf{A}_\Delta)^\infty$. Formulae (25),(26) remain valid for the mixture $S$, as well.

All the measures $\hat{P}_\gamma^{S\mathcal{M}}$ considered above have important common property coming from the equation (26):

$$\begin{aligned} & \hat{P}_\gamma^{S\mathcal{M}}(X_n \in \Gamma_{\mathbf{X}}|X_0, A_1, \ldots, X_{n-1}, A_n,) \\ = \ & I\{X_{n-1} \neq \Delta\} \frac{q(\Gamma_{\mathbf{X}} \setminus \{X_{n-1}\}|X_{n-1}, A_n)}{q_{X_{n-1}}(A_n)} + I\{q_{X_{n-1}}(A_n) = 0 \ or \ X_{n-1} = \Delta\} I\{\Gamma_{\mathbf{X}} \ni \Delta\}, \end{aligned}$$

meaning that all of them are strategic measures in the discrete-time Markov Decision Process $\mathcal{M}$ with state and action spaces $\mathbf{X}_\Delta$ and $\mathbf{A}_\Delta$ and with the transition probability

$$Q(y \in \Gamma_{\mathbf{X}}|x, a) = \begin{cases} \frac{q(\Gamma_{\mathbf{X}} \setminus \{x\}|x,a)}{q_x(a)}, & \text{if } x \neq \Delta,\ q_x(a) \neq 0; \\ I\{\Gamma_{\mathbf{X}} \ni \Delta\} & \text{otherwise.} \end{cases}$$

[4, Ch.3,§5].

As is known [18, Lemma 2], there exists a sequence of stochastic kernels $p_n^M(da_n|x_{n-1})$, $n = 1, 2, \ldots$, i.e. a Markov strategy in $\mathcal{M}$, defining a Markov standard $\xi$-strategy $S^M$, such that

$$\hat{E}_\gamma^{S\mathcal{M}}[I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} I\{A_n \in \Gamma_{\mathbf{A}}\}] = \hat{E}_\gamma^{S^M\mathcal{M}}[I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} I\{A_n \in \Gamma_{\mathbf{A}}\}], \quad n = 1, 2, \ldots$$

for all $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X}_\Delta)$, $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A}_\Delta)$. Since formula (25) is strategy-independent, we conclude that $\eta_n^S = \eta_n^{S^M}$, $n = 1, 2, \ldots$ and $\mathcal{D}_{st} \subset \mathcal{D}_\xi$.

Now, show that $\mathcal{D}_\xi \subset \mathcal{D}_{dm}$. Let $S^M = \{\mathbf{A},\ p_n^M,\ n = 1, 2, \ldots\}$ be a Markov standard $\xi$-strategy. It is known that the strategic measure $\hat{P}_\gamma^{S^M\mathcal{M}}$ in $\mathcal{M}$ (generated by a Markov strategy $p_n^M$) can be represented as

$$\hat{P}_\gamma^{S^M\mathcal{M}} = \int_{\mathbf{\Xi}^0} \xi_0^0 \hat{p}_0(d\xi_0^0), \tag{27}$$

where $\mathbf{\Xi^0}$, defined as

$$\mathbf{\Xi}^0 = \{\hat{P}_\gamma^{S\mathcal{M}}, S = \{\hat{\varphi}_n,\ n = 1, 2, \ldots\} \text{ are all possible simple deterministic Markov strategies in } \mathcal{M}\}, \tag{28}$$

is a Borel space, and $\hat{p}_0$ is a probability measure on $\mathbf{\Xi}^0$. For more details see [6, sections 2,3; Th.5.2].

For a fixed $\xi_0^0 \in \mathbf{\Xi}^0$ and $n = 1, 2, \ldots$, let $\xi_0^{0n}$ be the marginal of the measure $\xi_0^0$:

$$\xi_0^{0n}(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = \xi_0^0((\mathbf{X}_\Delta \times \mathbf{A}_\Delta)^{n-1} \times \Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}} \times (\mathbf{X}_\Delta \times \mathbf{A}_\Delta)^\infty),$$

$\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X}_\Delta)$, $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A}_\Delta)$. The mapping $\xi_0^{0n} = f^n(\xi_0^0)$ is measurable and even continuous if we fix the corresponding topologies in the state and action spaces and the weak topologies in the probability measures spaces. Using Corollary 7.27.1 [1], we see that, for stochastic kernel $k(dx, da|\xi_0^{0n}) \stackrel{\triangle}{=} \xi_0^{0n}(dx, da)$, there are measurable stochastic kernels $k_A(\Gamma_{\mathbf{A}}|x, \xi_0^{0n})$ and $k_X(\Gamma_{\mathbf{X}}|\xi_0^{0n}) = \xi_0^{0n}(\Gamma_{\mathbf{X}} \times \mathbf{A}_\Delta)$ on $\mathbf{A}_\Delta$ and $\mathbf{X}_\Delta$ respectively, such that

$$\xi_0^{0n}(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = \int_{\Gamma_{\mathbf{X}}} k_A(\Gamma_{\mathbf{A}}|x, \xi_0^{0n}) k_X(dx|\xi_0^{0n}).$$

Consider the mixture $S = \{\mathbf{\Xi}^0 \times \mathbf{A},\ \hat{p}_0,\ \hat{p}_n,\ n = 1, 2, \ldots\}$ of standard $\xi$-strategies $S^{\xi_0^0}$, where $\hat{p}_n(da_n|\xi_0^0, x_{n-1}) \stackrel{\triangle}{=} k_A(da_n|x_{n-1}, f^n(\xi_0^0))$ (see Definition 5) and prove that it is a mixture of simple deterministic Markov strategies. Since $\xi_0^0 = \hat{P}_\gamma^{S(\xi_0^0)\mathcal{M}}$ is a strategic measure in the Markov Decision Process $\mathcal{M}$ for some (simple) deterministic Markov strategy $S(\xi_0^0) = \{\hat{\varphi}_n^{\xi_0^0},\ n = 1, 2, \ldots\}$,

$$k_A(\Gamma_{\mathbf{A}}|x, \xi_0^{0n}) = I\{\Gamma_{\mathbf{A}} \ni \hat{\varphi}_n^{\xi_0^0}(x)\}$$

for $\xi_0^{0n}(dx \times \mathbf{A}_\Delta)$-almost all $x \in \mathbf{X}_\Delta$. Equivalently,

$$\hat{p}_n(\Gamma_{\mathbf{A}}|\xi_0^0, X_{n-1}) = I\{\Gamma_{\mathbf{A}} \ni \hat{\varphi}_n^{\xi_0^0}(X_{n-1})\} \qquad \hat{P}_\gamma^{S(\xi_0^0)\mathcal{M}}\text{-a.s.} \quad n = 1, 2, \ldots$$

The induction argument, when $n = 1, 2, \ldots$, implies that (for a fixed $\xi_0^0 \in \mathbf{\Xi}^0$), for the Markov strategy $S^{\xi_0^0} \stackrel{\triangle}{=} \{\hat{p}_n(da_n|\xi_0^0, x_{n-1}),\ n = 1, 2, \ldots\}$ in $\mathcal{M}$, equality $\hat{P}_\gamma^{S^{\xi_0^0}\mathcal{M}} = \hat{P}_\gamma^{S(\xi_0^0)\mathcal{M}}$ is valid. Here, with some abuse of notation, $S^{\xi_0^0}$ is a Markov strategy in $\mathcal{M}$ and also a Markov standard $\xi$-strategy in the original model. We proved that

$$\hat{p}_n(\Gamma_{\mathbf{A}}|\xi_0^0, X_{n-1}) = I\{\Gamma_{\mathbf{A}} \ni \hat{\varphi}_n^{\xi_0^0}(X_{n-1})\} \qquad \hat{P}_\gamma^{S^{\xi_0^0}\mathcal{M}}\text{-a.s.} \quad n = 1, 2, \ldots \tag{29}$$

As was mentioned above, when returning back to the continuous-time model, the measures $\hat{P}_\gamma^{S^{\xi_0^0}\mathcal{M}} = \hat{P}_\gamma^{S(\xi_0^0)\mathcal{M}}$ give rise to the measures $\hat{P}_\gamma^{S^{\xi_0^0}} = \hat{P}_\gamma^{S(\xi_0^0)}$ on $\hat{\Omega}$ (22), simply by applying formula (25). Now, the equality (29) holds $\hat{P}_\gamma^{S^{\xi_0^0}}$-a.s. and hence $P_\gamma^S$-a.s. because the strategic measure $P_\gamma^S$ has the form $P_\gamma^S(d\xi_0^0, d\hat{\omega}) = \hat{p}_0(d\xi_0^0)\hat{P}_\gamma^{S^{\xi_0^0}}(d\hat{\omega})$.

Thus $S$ is a mixture of simple deterministic Markov strategies $\{\hat{\varphi}_n^{\xi_0^0},\ n = 1, 2, \ldots\} = S(\xi_0^0)$.

Formula (27) implies that

$$P_\gamma^{S^M}(d\hat{\omega}) = \int_{\mathbf{\Xi}^0} \hat{p}_0(d\xi_0^0) P_\gamma^{S(\xi_0^0)}(d\hat{\omega}) = \int_{\mathbf{\Xi}^0} \hat{p}_0(d\xi_0^0) P_\gamma^{S^{\xi_0^0}}(d\hat{\omega}) = P_\gamma^S(\mathbf{\Xi}^0 \times d\hat{\omega}).$$

Hence $\eta_n^{S^M}(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = \eta_n^S(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}})$ for all $n = 1, 2, \ldots$

We proved that $\mathcal{D}_\xi \subset \mathcal{D}_{dm}$. Since $\mathcal{D}_{dm} \subset \mathcal{D}_{st} \subset \mathcal{D}_\xi$, the proof is completed. ■

Proof of Theorem 4. For a fixed $n = 1, 2, \ldots$,

$$
\begin{aligned}
& \eta_n^S(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) \\
= \; & E_\gamma^S \left[ E_\gamma^S \left[ I\{X_{n-1} \in \mathbf{X}\} \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | H_{n-1}, \Xi_n, t - T_{n-1}) dt | H_{n-1} \right] \right] \\
= \; & E_\gamma^S \left[ I\{X_{n-1} \in \mathbf{X}\} \int_{\mathbf{\Xi}_\Delta} p_n(d\xi | H_{n-1}) \left[ \int_{(0,\infty)} \int_{(0,\theta]} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | H_{n-1}, \xi, u) du \right. \right. \\
& \left. \left. \times \; \lambda_n(\mathbf{X}_\Delta | H_{n-1}, \xi, \theta) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \xi, \theta)} \, d\theta \right] \right] \quad \text{(change the order)} \\
= \; & E_\gamma^S \left[ I\{X_{n-1} \in \mathbf{X}\} \int_{\mathbf{\Xi}_\Delta} p_n(d\xi | H_{n-1}) \int_{(0,\infty)} \left( \int_{[s,\infty)} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | H_{n-1}, \xi, s) \right. \right. \\
& \left. \left. \times \; \lambda_n(\mathbf{X}_\Delta | H_{n-1}, \xi, \theta) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \xi, \theta)} \, d\theta \right) ds \right] \\
= \; & E_\gamma^S \left[ I\{X_{n-1} \in \mathbf{X}\} \int_{\mathbf{\Xi}_\Delta} p_n(d\xi | H_{n-1}) \int_{(0,\infty)} g(H_{n-1}, s) e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \xi, s)} \, ds \right],
\end{aligned}
$$

where, under fixed $\xi, \Gamma_{\mathbf{A}}, \Gamma_{\mathbf{X}}$, function $g$ is defined as $g(h_{n-1}, s) \stackrel{\triangle}{=} I\{x_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | h_{n-1}, \xi, s)$.

The last integral can be evaluated, after we notice that

$$
\begin{aligned}
e^{-\Lambda_n(\mathbf{X}_\Delta, h_{n-1}, \xi, s)} \; = \; & e^{-\int_{(0,s)} \int_{\mathbf{A}} \pi_n(da | h_{n-1}, \xi, u) \alpha(x_{n-1}, a) du} \\
& \times \left[ \int_{(s,\infty)} \hat{\lambda}(\mathbf{X} | h_{n-1}, \xi, v) e^{-\hat{\Lambda}_n(\mathbf{X}, h_{n-1}, \xi, v)} \, dv + e^{-\hat{\Lambda}_n(\mathbf{X}, h_{n-1}, \xi, \infty)} \right],
\end{aligned}
$$

in the following way:

$$
\begin{aligned}
& \int_{(0,\infty)} g(h_{n-1}, s) e^{-\Lambda_n(\mathbf{X}_\Delta, h_{n-1}, \xi, s)} \, ds \quad \text{(change the order)} \\
= \; & \int_{(0,\infty)} \int_{(0,v)} g(h_{n-1}, s) e^{-\int_{(0,s)} \int_{\mathbf{A}} \pi_n(da | h_{n-1}, \xi, u) \alpha(x_{n-1}, a) du} \hat{\lambda}(\mathbf{X} | h_{n-1}, \xi, v) e^{-\hat{\Lambda}_n(\mathbf{X}, h_{n-1}, \xi, v)} \, ds \, dv \\
& + \int_{(0,\infty)} g(h_{n-1}, s) e^{-\int_{(0,s)} \int_{\mathbf{A}} \pi_n(da | h_{n-1}, \xi, u) \alpha(x_{n-1}, a) du} \, ds \times e^{-\hat{\Lambda}_n(\mathbf{X}, h_{n-1}, \xi, \infty)}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\eta_n^S(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) \; = \; & E_\gamma^S \left[ I\{X_{n-1} \in \mathbf{X}\} \right. \\
& \times \int_{\bar{\mathbb{R}}_+ \times \mathbf{\Xi}_\Delta \times \mathbf{X}_\Delta} \hat{G}_n(d\theta, d\xi, dx | H_{n-1}) \left\{ \int_{(0,\theta] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | H_{n-1}, \xi, v) \right. \\
& \left. \left. \times e^{-\int_{(T_{n-1}, T_{n-1}+v]} \int_{\mathbf{A}} \pi_n(da | H_{n-1}, \xi, w - T_{n-1}) \alpha(X_{n-1}, a) dw} dv \right\} \right] \\
= \; & E_\gamma^S \left[ I\{X_{n-1} \in \mathbf{X}\} \int_{\bar{\mathbb{R}}_+ \times \mathbf{\Xi}_\Delta \times \mathbf{X}_\Delta} \hat{G}_n(d\theta, d\xi, dx | H_{n-1}) \right. \\
& \times \left\{ \int_{(T_{n-1}, T_{n-1}+\theta] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}} | H_{n-1}, \xi, t - T_{n-1}) \right. \\
& \left. \left. \times e^{-\int_{(T_{n-1}, t]} \int_{\mathbf{A}} \pi_n(da | H_{n-1}, \xi, w - T_{n-1}) \alpha(X_{n-1}, a) dw} dt \right\} \right].
\end{aligned}
$$

The last expression has the form $E^S_\gamma[I\{X_{n-1}\in\mathbf{X}\}\cdot F(H_{n-1})]$. Applying the similar, but simpler calculations, we obtain

$$E^S_\gamma[E^S_\gamma[I\{X_{n-1}\in\mathbf{X}\}\cdot F(H_{n-2},\Xi_{n-1},\Theta_{n-1},X_{n-1})|H_{n-2}]]$$

$$\begin{aligned}
&= E^S_\gamma\Bigg[I\{X_{n-2}\in\mathbf{X}\}\cdot\int_{\mathbf{\Xi}_\Delta}p_{n-1}(d\xi|H_{n-2})\int_{(0,\infty)}\int_{\mathbf{X}}I\{x\in\mathbf{X}\}F(H_{n-2},\xi,\theta,x)\\
&\quad\times\lambda_{n-1}(dx|H_{n-2},\xi,\theta)e^{-\Lambda_{n-1}(\mathbf{X}_\Delta,H_{n-2},\xi,\theta)}d\theta\Bigg]\\
&= E^S_\gamma\Bigg[I\{X_{n-2}\in\mathbf{X}\}\int_{\bar{\mathbb{R}}_+\times\mathbf{\Xi}_\Delta\times\mathbf{X}_\Delta}\hat{G}_{n-1}(d\theta,d\xi,dx|H_{n-2})\\
&\quad\times e^{-\int_{(0,\theta]}\int_{\mathbf{A}}\pi_{n-1}(da|H_{n-2},\xi,u)\alpha(X_{n-2},a)du}\,I\{x\in\mathbf{X}\}F(H_{n-2},\xi,\theta,x)\Bigg]\\
&= E^S_\gamma\Bigg[I\{X_{n-2}\in\mathbf{X}\}\int_{\bar{\mathbb{R}}_+\times\mathbf{\Xi}_\Delta\times\mathbf{X}_\Delta}\hat{G}_{n-1}(d\theta,d\xi,dx|H_{n-2})\\
&\quad\times\int_{\bar{\mathbb{R}}_+\times\mathbf{\Xi}_\Delta\times\mathbf{X}_\Delta}\hat{G}_n(d\tilde{\theta},d\tilde{\xi},d\tilde{x}|H_{n-2},\xi,\theta,x)e^{-\int_{(T_{n-2},T_{n-2}+\theta]}\int_{\mathbf{A}}\pi_{n-1}(da|H_{n-2},\xi,w-T_{n-2})\alpha(X_{n-2},a)dw}\\
&\quad\times\Bigg\{\int_{(T_{n-2}+\theta,T_{n-2}+\theta+\tilde{\theta}]\cap\mathbb{R}_+}I\{x\in\Gamma_{\mathbf{X}}\}\pi_n(\Gamma_{\mathbf{A}}|H_{n-2},\xi,\theta,x,\tilde{\xi},t-T_{n-2}-\theta)\\
&\quad\times\ e^{-\int_{(T_{n-2}+\theta,t]}\int_{\mathbf{A}}\pi_n(da|H_{n-2},\xi,\theta,x,\tilde{\xi},w-T_{n-2}-\theta)\alpha(x,a)dw}dt\Bigg\}\Bigg].
\end{aligned}$$

Continuing in the same way, we obtain the desired expression. ■

Proof of Theorem 5. Fix an arbitrary $\varepsilon>0$. We intend to provide the explicit formulae for $\tilde{p}_{n,k}$. For a fixed $n\geq 1$, we introduce random functions $Q_k(w)$ depending on $\omega\in\Omega$:

$$Q_k(w)\stackrel{\triangle}{=}\frac{\varepsilon(\varepsilon w)^{k-1}}{(k-1)!}e^{-\varepsilon w-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w)},\quad k=1,2,\ldots,\ w\in\mathbb{R}^0_+$$

and (random) function $f_w(t)$:

$$f_w(t)\stackrel{\triangle}{=}[\lambda_n(\mathbf{X}_\Delta|H_{n-1},\Xi_n,w+t)+\varepsilon]e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w+t)+\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w)-\varepsilon t},\qquad w,t\in\mathbb{R}^0_+.$$

The Poisson-related $\xi$-strategy $S^P$ under consideration is defined by

$$\tilde{p}_{n,0}(\Gamma_{\mathbf{A}}|x_{n-1})\stackrel{\triangle}{=}E^S_\gamma\left[\int_{(0,\infty)}f_0(t)\int_{(0,t]}\int_{\Gamma_{\mathbf{A}}}\pi_n(da|H_{n-1},\Xi_n,u)[q_{X_{n-1}}(a)+\varepsilon]du\ dt|X_{n-1}=x_{n-1}\right];$$

$$\tilde{p}_{n,k}(\Gamma_{\mathbf{A}}|x_{n-1})\stackrel{\triangle}{=}$$

$$\frac{E^S_\gamma\left[\int\limits_{(0,\infty)}Q_k(w)\int\limits_{(0,\infty)}f_w(t)\int\limits_{(0,t]}\int\limits_{\Gamma_{\mathbf{A}}}\pi_n(da|H_{n-1},\Xi_n,w+u)[q_{X_{n-1}}(a)+\varepsilon]dudtdw|X_{n-1}=x_{n-1}\right]}{E^S_\gamma\left[\int\limits_{(0,\infty)}Q_k(w)dw|X_{n-1}=x_{n-1}\right]},$$

for $k\geq 1$, and we plan to prove that $\eta^S_n=\eta^{S^P}_n$.

Below, $Z_k$ is the independent of anything RV having the $Erlang(\varepsilon,k)$ distribution. Clearly, under the control strategy $S$, the conditional probability $P^S_\gamma(Z_k<\Theta_n|X_{n-1}=x_{n-1})$ equals

$E_\gamma^S\left[\int_{(0,\infty)} Q_k(w)dw|X_{n-1}=x_{n-1}\right]$. Similarly, $P_\gamma^{S^P}(Z_k<\Theta_n|X_{n-1}=x_{n-1})=\prod_{i=1}^k p_i$, where $p_i=\int_{\mathbf{A}}\int_{(0,\infty)}\varepsilon e^{-\varepsilon w}e^{-q_{x_{n-1}}(a)w}dw\ \tilde{p}_{n,i-1}(da|x_{n-1})$, and we are going to prove by induction that these two probabilities coincide:

$$P_\gamma^{S^P}(Z_k<\Theta_n|X_{n-1}=x_{n-1})=P_\gamma^S(Z_k<\Theta_n|X_{n-1}=x_{n-1})=E_\gamma^S\left[\int_{(0,\infty)}Q_k(w)dw|X_{n-1}=x_{n-1}\right] \tag{30}$$

Below, in the case of the $S^P$ strategy, $\sum_{i=1}^k\tau_i^n$ usually plays the role of $Z_k$.

If $k=1$ then

$$\begin{aligned}p_1 &= \int_{\mathbf{A}}\int_{(0,\infty)}\varepsilon e^{-\varepsilon w}e^{-q_{x_{n-1}}(a)w}dwE_\gamma^S\left[\int_{(0,\infty)}[\lambda_n(\mathbf{X}_\Delta|H_{n-1},\Xi_n,t)+\varepsilon]e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,t)-\varepsilon t}\right.\\ &\left.\times\int_{(0,t]}[q_{X_{n-1}}(a)+\varepsilon]\pi_n(da|H_{n-1},\Xi_n,u)du\ dt|X_{n-1}=x_{n-1}\right].\end{aligned}$$

We move $[q_{x_{n-1}}(a)+\varepsilon]$ outside the conditional mathematical expectation and integrate w.r.t. $w$: $\int_{(0,\infty)}e^{-\varepsilon w-q_{x_{n-1}}(a)w}[q_{x_{n-1}}(a)+\varepsilon]dw=1$. Here and below, we use the Fubini theorem without special remarks. After integrating the result by parts w.r.t. $t$, we obtain:

$$p_1=E_\gamma^S\left[\int_{(0,\infty)}\varepsilon e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,t)-\varepsilon t}dt|X_{n-1}=x_{n-1}\right]=E_\gamma^S\left[\int_{(0,\infty)}Q_1(w)dw|X_{n-1}=x_{n-1}\right].$$

Suppose $\prod_{i=1}^k p_i=E_\gamma^S\left[\int_{(0,\infty)}Q_k(w)dw|X_{n-1}=x_{n-1}\right]$ for some $k\geq 1$ and prove the same equality for $k+1$ using (30).

$$\begin{aligned}\prod_{i=1}^{k+1}p_i &= E_\gamma^S\left[\int_{\mathbf{A}}\int_{(0,\infty)}\varepsilon e^{-\varepsilon v-q_{X_{n-1}}(a)v}dv\int_{(0,\infty)}Q_k(w)\int_{(0,\infty)}[\lambda_n(\mathbf{X}_\Delta|H_{n-1},\Xi_n,w+t)+\varepsilon]\right.\\ &\quad\times e^{-\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w+t)+\Lambda_n(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w)-\varepsilon t}\\ &\quad\left.\times\int_{(0,t]}\pi_n(da|H_{n-1},\Xi_n,w+t)[q_{X_{n-1}}+\varepsilon]du\ dt\ dw|X_{n-1}=x_{n-1}\right]\\ &= E_\gamma^S\left[\int_{(0,\infty)}\int_{(0,\infty)}\varepsilon t\frac{\varepsilon(\varepsilon w)^{k-1}}{(k-1)!}e^{-\varepsilon w-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w+t)-\varepsilon t}\right.\\ &\quad\left.\times[\lambda_n(\mathbf{X}_\Delta|H_{n-1},\Xi_n,w+t)+\varepsilon]dt\ dw|X_{n-1}=x_{n-1}\right]\quad(\text{denote } s=w+t)\\ &= E_\gamma^S\left[\varepsilon\int_{(0,\infty)}\frac{\varepsilon(\varepsilon w)^{k-1}}{(k-1)!}\left\{\int_{(w,\infty)}s[\lambda_n(\mathbf{X}_\Delta|H_{n-1},\Xi_n,s)+\varepsilon]e^{-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,s)-\varepsilon s}ds\right.\right.\\ &\quad\left.\left.-\ we^{-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,w)}\right\}dw|X_{n-1}=x_{n-1}\right]\quad(\text{integration by parts w.r.t. } s)\\ &= E_\gamma^S\left[\varepsilon\int_{(0,\infty)}\frac{\varepsilon(\varepsilon w)^{k-1}}{(k-1)!}\int_{(w,\infty)}e^{-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,s)-\varepsilon s}ds\ dw|X_{n-1}=x_{n-1}\right]\\ &= E_\gamma^S\left[\varepsilon\int_{(0,\infty)}\int_{(0,s)}\frac{\varepsilon(\varepsilon w)^{k-1}}{(k-1)!}e^{-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,s)-\varepsilon s}dw\ ds|X_{n-1}=x_{n-1}\right]\\ &= E_\gamma^S\left[\varepsilon\int_{(0,\infty)}\frac{(\varepsilon s)^k}{(k)!}e^{-\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,s)-\varepsilon s}\ ds|X_{n-1}=x_{n-1}\right],\end{aligned}$$

what we wanted to prove.

The next step is to prove that

$$P_\gamma^{S^P}(X_n \in \Gamma_{\mathbf{X}}) = P_\gamma^S(X_n \in \Gamma_{\mathbf{X}}) \tag{31}$$

for all $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$, $n = 0, 1, 2, \ldots$. This equality is obviously valid at $n = 0$ because the initial distribution $\gamma$ is fixed. Suppose it holds for some $n - 1 \geq 0$ and prove that

$$P_\gamma^{S^P}(X_n \in \Gamma_{\mathbf{X}} | X_{n-1} = x_{n-1}) = P_\gamma^S(X_n \in \Gamma_{\mathbf{X}} | X_{n-1} = x_{n-1}). \tag{32}$$

Clearly, it is sufficient to consider the case $\Theta_n < \infty$. Using (30) and the property $\lim_{k\to\infty} \sum_{i=0}^k \tau_i^n = \infty$ $P_\gamma^{S^P}$-a.s., we obtain

$$\begin{aligned}
& P_\gamma^{S^P}(X_n \in \Gamma_{\mathbf{X}} | X_{n-1} = x_{n-1}) \\
= & \sum_{k=0}^\infty P_\gamma^{S^P}(X_n \in \Gamma_{\mathbf{X}},\ \sum_{i=0}^k \tau_i^n \leq \Theta_n < \sum_{i=0}^{k+1} \tau_i^n | X_{n-1} = x_{n-1}) = E_\gamma^S \left[ \int_{\mathbf{A}} \int_{(0,\infty)} f_0(t) \right. \\
& \left. \times \int_{(0,t]} \pi_n(da|H_{n-1}, \Xi_n, u)[q_{X_{n-1}}(a) + \varepsilon] du\ dt\ \frac{q(\Gamma_{\mathbf{X}} \setminus \{X_{n-1}\}|X_{n-1}, a)}{q_{X_{n-1}}(a) + \varepsilon} | X_{n-1} = x_{n-1} \right] \\
& + \sum_{k=1}^\infty E_\gamma^S \left[ \int_{\mathbf{A}} \int_{(0,\infty)} Q_k(w) \int_{(0,\infty)} f_w(t) \int_{(0,t]} \pi_n(da|H_{n-1}, \Xi_n, w + u) \right. \\
& \left. \times [q_{X_{n-1}}(a) + \varepsilon] du\ dt\ dw\ \frac{q(\Gamma_{\mathbf{X}} \setminus \{X_{n-1}\}|X_{n-1}, a)}{q_{X_{n-1}}(a) + \varepsilon} | X_{n-1} = x_{n-1} \right] \\
= & E_\gamma^S \left[ \int_{(0,\infty)} f_0(t) \int_{(0,t]} \lambda_n(\Gamma_{\mathbf{X}}|H_{n-1}, \Xi_n, u) du\ dt | X_{n-1} = x_{n-1} \right] \\
& + \sum_{k=1}^\infty E_\gamma^S \left[ \int_{(0,\infty)} Q_k(w) \int_{(0,\infty)} f_w(t) \int_{(0,t]} \lambda_n(\Gamma_{\mathbf{X}}|H_{n-1}, \Xi_n, w + u) du\ dt\ dw | X_{n-1} = x_{n-1} \right] \\
= & E_\gamma^S \left[ \int_{(0,\infty)} e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, t) - \varepsilon t} \lambda_n(\Gamma_{\mathbf{X}}|H_{n-1}, \Xi_n, t) dt | X_{n-1} = x_{n-1} \right] \\
& + \sum_{k=1}^\infty E_\gamma^S \left[ \int_{(0,\infty)} Q_k(w) \int_{(0,\infty)} e^{-\Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, w+t) + \Lambda_n(\mathbf{X}_\Delta, H_{n-1}, \Xi_n, w) - \varepsilon t} \right. \\
& \left. \times \lambda_n(\Gamma_{\mathbf{X}}|H_{n-1}, \Xi_n, w + t) dt\ dw | X_{n-1} = x_{n-1} \right] \\
= & E_\gamma^S \left[ \int_{(0,\infty)} f_0(t) \frac{\lambda_n(\Gamma_{\mathbf{X}}|H_{n-1}, \Xi_n, t)}{\lambda_n(\mathbf{X}_\Delta|H_{n-1}, \Xi_n, t) + \varepsilon} dt | X_{n-1} = x_{n-1} \right] \\
& + \sum_{k=1}^\infty E_\gamma^S \left[ \int_{(0,\infty)} Q_k(w) \int_{(0,\infty)} f_w(t) \frac{\lambda_n(\Gamma_{\mathbf{X}}|H_{n-1}, \Xi_n, w + t)}{\lambda_n(\mathbf{X}_\Delta|H_{n-1}, \Xi_n, w + t) + \varepsilon} dt\ dw | X_{n-1} = x_{n-1} \right] \\
= & \sum_{k=0}^\infty P_\gamma^S(X_n \in \Gamma_{\mathbf{X}},\ \sum_{i=0}^k \hat{\tau}_i^n \leq \Theta_n < \sum_{i=0}^{k+1} \hat{\tau}_i^n | X_{n-1} = x_{n-1}),
\end{aligned}$$

where $\hat{\tau}_0 = 0$ and $\{\hat{\tau}_i\}_{i=1}^\infty$ is a sequence of independent $Exp(\varepsilon)$ RVs. Formulae (32) and hence (31) are proved.

Although the occupation measures may be not finite, formula

$$\eta_n^S(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = E_\gamma^S \left[ E_\gamma^S \left[ \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}}|H_{n-1}, \Xi_n, t - T_{n-1}) dt | X_{n-1} \right] \right]$$

(and the similar formula for $S^P$) is valid [17, §IV.3]. Therefore, to complete the proof of the

theorem, we need to establish equality

$$\begin{aligned} D^S(\Gamma_{\mathbf{A}}|x) &\stackrel{\triangle}{=} E_\gamma^S\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}\pi_n(\Gamma_{\mathbf{A}}|H_{n-1},\Xi_n,t-T_{n-1})dt|X_{n-1}=x\right] \\ &= D^{S^P}(\Gamma_{\mathbf{A}}|x)\stackrel{\triangle}{=}E_\gamma^{S^P}\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}p_n(d\xi_n|x)\times I\{\varphi_n(\xi_n,t-T_{n-1})\in\Gamma_{\mathbf{A}}\}dt\right], \end{aligned} \tag{33}$$

because $\forall\Gamma_{\mathbf{X}}\in\mathcal{B}(\mathbf{X})$

$$\eta_n^S(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})=\int_{\Gamma_{\mathbf{X}}}D^S(\Gamma_{\mathbf{A}}|x)P_\gamma^S(X_{n-1}\in dx);\quad \eta_n^{S^P}(\Gamma_{\mathbf{X}}\times\Gamma_{\mathbf{A}})=\int_{\Gamma_{\mathbf{X}}}D^{S^P}(\Gamma_{\mathbf{A}}|x)P_\gamma^{S^P}(X_{n-1}\in dx)$$

and the distributions of $X_{n-1}$ under the control strategies $S$ and $S^P$ coincide. Here and below, the set $\Gamma_{\mathbf{A}}\in\mathcal{B}(\mathbf{A})$ is arbitrarily fixed.

Using (30), we obtain

$$\begin{aligned} & D^{S^P}(\Gamma_{\mathbf{A}}|x) \\ = & \int_{\Gamma_{\mathbf{A}}}\tilde{p}_{n,0}(da|x)\frac{1}{q_x(a)+\varepsilon} \\ & +\sum_{k=1}^\infty E_\gamma^S\left[\int_{(0,\infty)}Q_k(w)dw|X_{n-1}=x\right]\int_{\Gamma_{\mathbf{A}}}\tilde{p}_{n,k}(da|x)\frac{1}{q_x(a)+\varepsilon} \\ = & E_\gamma^S\left[\int_{(0,\infty)}f_0(t)\int_{(0,t]}\pi_n(da|H_{n-1},\Xi_n,u)du\ dt|X_{n-1}=x\right] \\ & +\sum_{k=1}^\infty E_\gamma^S\left[\int_{(0,\infty)}Q_k(w)\int_{(0,\infty)}f_w(t)\int_{(0,t]}\pi_n(da|H_{n-1},\Xi_n,w+u)du\ dt\ dw|X_{n-1}=x\right]. \end{aligned}$$

We evaluate the second term $\sum_{k=1}^\infty$ separately using the abbreviated notations

$$\lambda(t)\stackrel{\triangle}{=}\lambda_n(\mathbf{X}_\Delta|H_{n-1},\Xi_n,t),\ \Lambda(t)\stackrel{\triangle}{=}\Lambda(\mathbf{X}_\Delta,H_{n-1},\Xi_n,t),\ \text{and}\ \pi(t)\stackrel{\triangle}{=}\pi_n(\Gamma_{\mathbf{A}}|H_{n-1},\Xi_n,t):$$

$$\begin{aligned} & E_\gamma^S\left[\int_{(0,\infty)}\varepsilon\int_{(0,\infty)}[\lambda(w+t)+\varepsilon]e^{-\Lambda(w+t)-\varepsilon t}\int_{(w,w+t]}\pi(u)du\ dt\ dw|X_{n-1}=x\right] \\ & \text{(denote } y\stackrel{\triangle}{=}w+t \text{ and change the order of integration)} \\ = & E_\gamma^S\left[\int_{(0,\infty)}[\lambda(y)+\varepsilon]e^{-\Lambda(y)-\varepsilon y}\left[\int_{(0,y)}\varepsilon e^{\varepsilon w}\int_{(w,y]}\pi(u)du\ dw\right]dy|X_{n-1}=x\right] \\ & \text{(integration by parts w.r.t. } w) \\ = & E_\gamma^S\left[\int_{(0,\infty)}[\lambda(y)+\varepsilon]e^{-\Lambda(y)-\varepsilon y}\left[\int_{(0,y)}(e^{\varepsilon w}-1)\pi(w)dw\right]dy|X_{n-1}=x\right]. \end{aligned}$$

Now

$$\begin{aligned} & D^{S^P}(\Gamma_{\mathbf{A}}|x) \\ = & E_\gamma^S\left[\int_{(0,\infty)}[\lambda(y)+\varepsilon]e^{-\Lambda(y)-\varepsilon y}\int_{(0,y)}e^{\varepsilon w}\pi(w)dw\ dy|X_{n-1}=x\right] \\ & \text{(integration by parts w.r.t. } y) \\ = & E_\gamma^S\left[\lim_{Y\to\infty}\left\{\int_{(0,Y)}e^{-\Lambda(y)-\varepsilon y}\cdot e^{\varepsilon y}\pi(y)dy-e^{-\Lambda(Y)-\varepsilon Y}\int_{(0,Y)}e^{\varepsilon w}\pi(w)dw\right\}|X_{n-1}=x\right]. \end{aligned}$$

Since

$$e^{-\varepsilon Y}\int_{(0,Y)} e^{\varepsilon w}\pi(w)dw \le \frac{1}{\varepsilon}(1-e^{-\varepsilon Y}) \le \frac{1}{\varepsilon}, \tag{34}$$

we conclude that

$$\lim_{Y\to\infty}\left\{\int_{(0,Y)} e^{-\Lambda(y)}\pi(y)dy - e^{-\Lambda(Y)-\varepsilon Y}\int_{(0,Y)} e^{\varepsilon w}\pi(w)dw\right\} = \int_{(0,\infty)} e^{-\Lambda(y)}\pi(y)dy \tag{35}$$

if the integral in the righthand side equals $+\infty$. Similarly, equality (35) holds true if $\lim_{Y\to\infty}\Lambda(Y) = \infty$ because of (34): $\lim_{Y\to\infty} e^{-\Lambda(Y)-\varepsilon Y}\int_{(0,Y)} e^{\varepsilon w}\pi(w)dw = 0$.

Suppose now that $\lim_{Y\to\infty}\Lambda(Y) < \infty$ and $\int_{(0,\infty)} e^{-\Lambda(y)}\pi(y)dy < \infty$. In this case, $\int_{(0,\infty)}\pi(y)dy < \infty$ and, for an arbitrarily fixed $\delta > 0$, we take $\hat{Y} \in (0,\infty)$ such that $\int_{(\hat{Y},\infty)}\pi(y)dy < \delta$. Now, considering only $Y > \hat{Y}$,

$$\overline{\lim_{Y\to\infty}}\left[e^{-\Lambda(Y)-\varepsilon Y}\int_{(0,\hat{Y}]} e^{\varepsilon w}\pi(w)dw + e^{-\Lambda(Y)-\varepsilon Y}\int_{(\hat{Y},Y)} e^{\varepsilon w}\pi(w)dw\right] \le \overline{\lim_{Y\to\infty}}\, e^{-\Lambda(Y)-\varepsilon Y}\delta e^{\varepsilon Y}$$

because

$$\lim_{Y\to\infty} e^{-\Lambda(Y)-\varepsilon Y}\int_{(0,\hat{Y}]} e^{\varepsilon w}\pi(w)dw = 0$$

and

$$\int_{(\hat{Y},Y)} e^{\varepsilon w}\pi(w)dw \le e^{\varepsilon Y}\int_{(\hat{Y},Y)}\pi(w)dw \le \delta e^{\varepsilon Y}.$$

Since $\delta > 0$ was arbitrary, in this case $\lim_{Y\to\infty} e^{-\Lambda(Y)-\varepsilon Y}\int_{(0,Y)} e^{\varepsilon w}\pi(w)dw = 0$.

Therefore, in any case we have equality (35) and

$$\begin{aligned} D^{S^P}(\Gamma_{\mathbf{A}}|x) &= E_\gamma^S\left[\int_{(0,\infty)} e^{-\Lambda(y)}\pi(y)dy|X_{n-1}=x\right] \quad \text{(integration by parts)} \\ &= E_\gamma^S\left[\int_{(0,\infty)}\lambda(y)e^{-\Lambda(y)}\int_{(0,y)}\pi(u)du\; dy + e^{-\Lambda(\infty)}\int_{(0,\infty)}\pi(y)dy|X_{n-1}=x\right] \\ &= D^S(\Gamma_{\mathbf{A}}|x). \end{aligned}$$

■

# References


[1] Bertsekas, D. and Shreve, S. *Stochastic Optimal Control.* Academic Press, NY, 1978.

[2] Bremaud, P. *Markov Chains: Gibbs Fields, Monte Carlo simulation, and Queues.* Springer, NY, 1999.

[3] Dellacherie, C. *Capacites et Processus Stochastiques.* Springer-Verlag, Berlin, 1972.

[4] Dynkin, E.B. and Yushkevich, A.A. *Controlled Markov Processes and their Applications.* Springer-Verlag, Berlin, 1979.

[5] Ethier, S.N. and Kurtz, T.G. *Markov Processes. Characterization and Convergence.* Wiley, NY, 1986.

[6] Feinberg, E.A.: On measurability and representation of strategic measures in Markov decision processes. In *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell, IMS Lecture Notes Monographs Ser.* (T.Ferguson ed.) **30** (1996) 29–43.

[7] Feinberg, E.: Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.* **29** (2004) 492-524.

[8] Ghosh, M. and Saha, S.: Non-stationary semi-Markov decision processes on a finite horizon. *Stoch. Anal. Appl.* **31** (2013) 183-190.

[9] Guo, X. and Zhang, Y.: Constrained total undiscounted continuous-time Markov decision processes. http://arxiv.org/pdf/1304.3314v5.pdf

[10] Guo, X. and Hernández-Lerma, O. *Continuous-Time Markov Decision Processes: Theory and Applications.* Springer-Verlag, Heidelberg, 2009.

[11] Guo, X. and Piunovskiy, A.: Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Oper. Res.* **36** (2011) 105-132.

[12] Hernández-Lerma, O. and Lasserre, J.B. *Discrete-Time Markov Control Processes.* Springer-Verlag, NY, 1996.

[13] Hernández-Lerma, O. and Lasserre, J.B. *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, NY, 1999.

[14] Huang, Y., Li, Z. and Guo, X.: Constrained optimality for finite horizon semi-Markov decision processes in Polish spaces. *Oper. Res. Lett.* **42** (2014) 123-129.

[15] Jacod, J.: Multivariate point processes: predictable projection, Radon-Nykodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebite.* **31** (1975) 235-253.

[16] Kitaev, M and Rykov, V. *Controlled Queueing Systems.* CRC Press, Boca Raton, 1995.

[17] Neveu, J. *Mathematical foundations of the calculus of probability.* Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam 1965.

[18] Piunovskiy, A. *Optimal Control of Random Sequences in Problems with Constraints.* Kluwer, Dordrecht, 1997.

[19] Piunovskiy, A. *Examples in Markov Decision Processes.* Imperial College Press, London, 2013.

[20] Piunovskiy, A. and Zhang, Y.: Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.*, **49** (2011) 2032-2061.

[21] Piunovskiy, A. and Zhang, Y.: The transformation method for continuous-time Markov decision processes. *J. Optim. Theory Appl.*, **154** (2012) 691-712.

[22] Piunovskiy, A.: Sufficient classes of strategies in continuous-time Markov decision processes with total expected cost. In *Modern Trends in Controlled Stochastic Processes* (A.Piunovskiy ed.) V.II (2016).

[23] Prieto-Rumeau, T. and Hernandez-Lerma, O. *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games.* Imperial College Press, London, 2012.

[24] Tijms, H.C. *A First Course in Stochastic Models.* Wiley, Chichester, 2003.

[25] Zhang, Y.: Average optimality for continuous-time Markov decision processes under weak continuity conditions. *J.Appl.Prob.*, **51** (2014) 954-970.